
Sistema automático de predicción estética basado en Computación Evolutiva y Deep Learning

NEREIDA RODRÍGUEZ FERNÁNDEZ

TESIS DOCTORAL 2022

Directores:

Dr. Juan Jesús Romero Cardalda

Dr. Adrián Carballal Mato

Programa oficial de doctorado en Tecnologías de la Información y las
Comunicaciones



UNIVERSIDADE DA CORUÑA



Dr. Juan Jesús Romero Cardalda, profesor titular en el área de Ciencias de la Computación e Inteligencia Artificial de la Universidade da Coruña y **Dr. Adrián Carballal Mato**, profesor contratado doctor en el área de Ciencias de la Computación e Inteligencia Artificial de la Universidade da Coruña,

HACEN CONSTAR QUE:

La tesis titulada ***Sistema automático de predicción estética basado en Computación Evolutiva y Deep Learning***, realizada por Dña. Nereida Rodríguez Fernández, con DNI 48114423X, bajo nuestra dirección en el Departamento de Ciencias de la Computación y Tecnologías de la información, cumple los requisitos para optar al grado de Doctor y obtener la mención internacional.

En A Coruña, a 5 de septiembre de 2022.

Fdo. Juan Jesús Romero Cardalda

Fdo. Adrián Carballal Mato



Código Seguro De Verificación	zFK4zwK6/8X3X/wqRPrdVQ==	Estado	Data e hora
Asinado Por	Juan Jesús Romero Cardalda	Asinado	05/09/2022 18:53:25
	Adrián Carballal Mato	Asinado	05/09/2022 09:24:27
Observacións		Páxina	1/1
Url De Verificación	https://sede.udc.gal/services/validation/zFK4zwK6/8X3X/wqRPrdVQ==		
Normativa	Este informe ten o carácter de copia electrónica auténtica con validez e eficacia administrativa de ORIGINAL (art. 27 Lei 39/2015).		



*A mis padres,
por apoyarme y abrazarme
en cada paso de mi vida.*

Agradecimientos

Gracias, en primer lugar, a mis directores, Juan y Adrián. Gracias por abrir una puerta que ni siquiera existía en mi camino, por confiar en mí y darme esta oportunidad. Gracias, especialmente, por escucharme y darme espacio para crecer, por dedicarme un tiempo que muchas veces no tenéis, y por guiarme con cariño durante los últimos años. Gracias por vuestra inestimable ayuda, por los consejos y por vuestra aportación al desarrollo de esta tesis en todas las formas posibles.

Gracias al grupo RNASA-IMEDIR y a todas las personas del laboratorio que me acompañaron en esta etapa. Gracias a Álvaro, a Iria, a Carlos y a todos aquellos que han aportado tiempo, conocimiento y trabajo a esta tesis.

Gracias a la vida por poner en este mismo camino a Sara y a Jose. Gracias Sara, por ser la mejor compañera en todos los sentidos posibles, por tu generosidad y sensibilidad, y por entenderme sin necesidad de hablar. A ti, Jose, gracias por ayudarme a ver siempre las cosas desde otro punto de vista, por tu franqueza, por los detalles, y por enseñarme a disfrutar de que todo fluya. Gracias a los dos por escucharme siempre y por estar a mi lado sin importar la distancia.

A mi familia, gracias por creer en mí y por hacerme llegar siempre su calor. Gracias a mis abuelos, mis tesoros, de los que tanto aprendí y sigo aprendiendo. A mi hermano, Fabián, por ese vínculo que solo él y yo podemos entender, por estar siempre en los momentos importantes y por quererme incondicionalmente.

Y, por supuesto, gracias a vosotros, mamá y papá, por educarme en los valores que hoy me han traído hasta aquí, por inculcarme el valor de la disciplina desde el amor y el respeto. Gracias por enseñarme a tomar mis propias decisiones y darme espacio para hacerlo. Gracias por estar siempre a mi lado, por acompañarme en los pasos inciertos y tener los brazos preparados tras los tropiezos.

Gracias a todos, de corazón.

*“La verdadera ciencia enseña,
por encima de todo, a dudar y a ser ignorante.”*

– Miguel de Unamuno

Resumen

Actualmente, con el auge de las redes sociales, acostumbramos a tomar decisiones en función del valor estético de las imágenes. En el comercio electrónico, por ejemplo, tomamos decisiones de compra en función de las imágenes del producto. En este contexto, un sistema automático que permita seleccionar y ordenar las imágenes en función de su valor estético puede ser de gran valor.

Esta tesis aborda diferentes problemas del campo de la estética computacional y propone nuevas soluciones que son validadas finalmente en un caso práctico real. En primer lugar, se estudian los datasets utilizados en estética computacional y se propone una nueva metodología para la creación de conjuntos de imágenes generalizables que se pueda aplicar en problemas de Machine Learning. A continuación, se expone un nuevo enfoque que utiliza *transfer learning* con un nuevo algoritmo genético híbrido para la predicción del valor estético en imágenes digitales. Finalmente, se aplica a un caso práctico real tanto la metodología de creación de datasets propuesta, como el modelo híbrido que ofreció mejores resultados en la fase experimental. Los resultados sugieren que utilizar estas herramientas en la vida cotidiana pueden mejorar tanto la experiencia de los usuarios como la productividad de los comercios electrónicos.

Resumo

Actualmente, co auxe das redes sociais, acostumamos a tomar decisións en función do valor estético das imaxes. No comercio electrónico, por exemplo, tomamos decisións de compra en función das imaxes do produto. Neste contexto, un sistema automático que permita seleccionar e ordenar as imaxes en función do seu valor estético pode ser de gran valor.

Esta tese aborda diferentes problemas do campo da estética computacional e propón novas solucións que son validadas finalmente nun caso práctico real. En primeiro lugar, estúdanse os datasets utilizados en estética computacional e proponse unha nova metodoloxía para a creación de conxuntos de imaxes xeneralizables que se poida aplicar en problemas de Machine Learning. A continuación, expónse un novo enfoque que utiliza transfer learning cun novo algoritmo xenético híbrido para a predición do valor estético en imaxes dixitais. Finalmente, aplícase a un caso práctico real tanto a metodoloxía de creación de datasets proposta, como o modelo híbrido que ofreceu mellores resultados na fase experimental. Os resultados suxiren que utilizar estas ferramentas na vida cotiá poden mellorar tanto a experiencia dos usuarios como a produtividade dos comercios electrónicos.

Abstract

Nowadays, with the rise of social media, we are used to making decisions based on the aesthetic value of images. In e-commerce, for example, we make purchasing decisions based on product images. In this context, an automatic system that allows us to select and sort images according to their aesthetic value can be of great value.

This thesis addresses different problems in the field of computational aesthetics and proposes new solutions that are finally validated in a real case study. First, we study the datasets used in computational aesthetics and propose a new methodology for the creation of generalisable image sets that can be applied to Machine Learning problems. Then, a new approach using transfer learning with a new hybrid genetic algorithm for the prediction of aesthetic value in digital images is presented. Finally, both the proposed dataset creation methodology and the hybrid model that gave the best results in the experimental phase are applied to a real case study. The results suggest that using these tools can improve both user experience and e-commerce productivity.

Capítulo 1

Prefacio

Desde hace miles de años, el ser humano ha perseguido la capacidad de entender el arte desde diferentes enfoques, tanto filosóficos como científicos. Con el desarrollo de la informática, pioneros como Ada Lovelace o Alan Turing llegaron a augurar la posibilidad de crear ordenadores con capacidades creativas y artísticas. Actualmente, campos como la inteligencia artificial o la psicología se han aproximado al estudio de la estética con enfoques independientes.

Como ejemplo significativo, en 2006, Datta et al. [35] presentan el primer dataset público para modelización estética, formado por 3.581 imágenes del portal de fotografía Photo.net y evaluadas en función de su valor estético y originalidad. A este trabajo le siguen otros con propuestas de datasets y sistemas basados en Machine Learning que constituyen el estado del arte actual en estética computacional. Sin embargo, existen muy pocos trabajos que validen los sistemas propuestos en casos prácticos reales.

Actualmente, vivimos en una realidad hiperconectada en la que recibimos miles de estímulos visuales a diario. Esto se acentúa cuando hablamos de Internet donde, especialmente con el auge de las redes sociales, consumimos una gran cantidad de contenido digital. Con ello, hemos aprendido a tomar decisiones en función de las imágenes que recibimos, ya sea para dar un *like* en Instagram o para decidir qué inmueble visitar cuando estamos buscando vivienda. Estas decisiones se toman, normalmente, en función de distintos factores y el valor estético de la imagen suele estar entre ellos. Así, un sistema automático que permita seleccionar y ordenar las imágenes en función de su estética puede ser de gran valor para la comunidad digital.

El grupo de investigación en Redes de Neuronas Artificiales y Sistemas Adap-

tativos (RNASA) de la Universidade da Coruña ha desarrollado diferentes experimentos que sirven de base al presente proyecto de tesis. Así, se realizaron (i) varios trabajos de clasificación de obras pictóricas según su autor, con resultados de acierto superiores al 90 % [22, 140], (ii) clasificación de imágenes según su valor estético empleando datasets procedentes de Photo.net y DPChallenge.com [139, 140], (iii) clasificación automática de imágenes entre fotografías y pinturas con un 94,82 % de acierto [21], y (iv) integración de un evaluador basado en redes de neuronas artificiales y un generador de programación genética en un sistema autónomo de creación de imágenes [108].

Esta tesis se inicia con el objetivo de corroborar que un modelo de Machine Learning capaz de medir el valor estético de imágenes digitales es útil en el comercio electrónico. Para ello, primero se abordan diferentes problemas de la estética computacional y se proponen nuevos enfoques que pueden mejorar los resultados del estado del arte. En primer lugar, se propone un nuevo dataset formado por imágenes del portal fotográfico DPChallenge y, con él, una nueva metodología para la creación de conjuntos de datos generalizables para el entrenamiento de modelos de Machine Learning. En segundo lugar, se utilizan dos CNNs diferentes para entrenar un algoritmo genético híbrido creado anteriormente por el grupo de investigación que respalda esta tesis doctoral y se comparan sus resultados con los del estado del arte. Finalmente, se valida en un caso práctico basado en el sector inmobiliario, tanto la metodología propuesta para la creación de datasets, como el modelo que ofrece mejores resultados en la fase experimental.

1.1. Objetivos

El principal objetivo de este trabajo de investigación es presentar un sistema de predicción estética basado en computación evolutiva y *deep learning* que permita aportar valor en aplicaciones comerciales. Para ello, se plantean los siguientes objetivos específicos:

1. Crear un nuevo dataset con mayor coherencia interna que los estudiados en el estado del arte, que sirva a la comunidad para el entrenamiento de modelos de Machine Learning dedicados a la medición de estética y calidad en imágenes digitales.
2. Presentar una nueva metodología para la creación de conjuntos de datos

generalizables que sea extrapolable a otros problemas de Machine Learning.

3. Comprobar que las métricas seleccionadas mediante *transfer learning* pueden aportar mejores resultados que las métricas ad-hoc en problemas de clasificación y predicción estética.
4. Probar un algoritmo genético híbrido, creado anteriormente por uno de los directores de esta tesis, con dos CNNs diferentes para clasificación y predicción estética y comparar sus resultados con los del estado del arte.
5. Validar en un ámbito específico, como es el inmobiliario, las propuestas planteadas en la fase experimental de este trabajo.

1.2. Metodología

A fin de alcanzar los objetivos expuestos en la sección anterior, se ha seguido una metodología basada en cinco etapas claramente diferenciadas, aunque no necesariamente separadas en el tiempo:

- Estudio del estado del arte en estética y estética computacional, desde un punto de vista filosófico, psicológico y científico, y de la predicción de su valor a través de sistemas informáticos de inteligencia artificial.
- Diseño y creación de diversos conjuntos de datos de entrenamiento para su uso en modelos basados en Machine Learning. Se recopilarán imágenes de diferentes fuentes públicas, algunas de entre las más utilizadas en el estado del arte, y se realizará un procesado de los datos y una evaluación de las imágenes por parte de una población determinada. La metodología específica para la creación de los conjuntos de datos se puede encontrar de forma más detallada en el capítulo 5.
- Experimentación con el sistema automático de predicción del valor estético. Las tareas de experimentación permiten validar las aproximaciones propuestas y compararlas con otros trabajos publicados. También se pretende comparar entre sí las diferentes propuestas y analizar en profundidad su relevancia para la tarea de predicción estética.

- Análisis de los resultados obtenidos. Se realizarán informes que indiquen las estimaciones estadísticas producidas en el conjunto de las ejecuciones que se lleven a cabo en el proceso de experimentación.
- Validación de las propuestas en una aplicación práctica. Para esto, se llevará a cabo una metodología específica explicada con detalle en el capítulo 7.

1.3. Estructura de la memoria

Esta tesis doctoral se estructura en nueve capítulos, organizados en tres grandes bloques. Tras este primer capítulo introductorio, comienza la Parte I, donde se hará un recorrido por el estado del arte a través de los capítulos 2, 3 y 4. A continuación, en la Parte II, se exponen las propuestas experimentales de esta tesis doctoral y los resultados alcanzados a través de los capítulos 5, 6 y 7. Por último, en la Parte III, se exponen las conclusiones generales obtenidas, junto con los futuros desarrollos previstos.

El capítulo 2 parte de las definiciones de estética y percepción visual para hacer un recorrido por la historia de la estética en campos como la filosofía y la ciencia, terminando con la exposición de distintas medidas del valor estético y pruebas psicológicas que permiten validar el juicio estético.

El capítulo 3 se centra en la estética computacional y en él se puede encontrar un análisis de las principales aportaciones computacionales que existen hasta el momento en el campo de la estética.

Para finalizar con el estado del arte, el capítulo 4 aborda el estudio de los diferentes conjuntos de datos utilizados hasta el momento en el campo de la estética computacional.

La Parte II comienza con el capítulo 5, donde se analizan algunas limitaciones de los datasets del estado del arte y se propone un nuevo conjunto de datos que pretende cubrir las necesidades detectadas en el campo. Además, se expone una nueva metodología para la construcción de conjuntos de datos extrapolable a otros problemas de Machine Learning.

El capítulo 6 expone un nuevo algoritmo genético híbrido para la predicción y clasificación estética de imágenes con *transfer learning*.

En el capítulo 7 se validan las propuestas de los dos capítulos anteriores en una aplicación práctica real con anuncios de portales y agencias inmobiliarias del parque nacional español.

Finalmente, los capítulos 8 y 9 exponen las conclusiones generales de este trabajo en castellano y en inglés, respectivamente, y plantean los desarrollos futuros que se pretenden llevar a cabo en el campo de la evaluación estética.

Índice general

Resumen	11
1. Prefacio	17
1.1. Objetivos	18
1.2. Metodología	19
1.3. Estructura de la memoria	20
I Estado del Arte	33
2. Estética y percepción	35
2.1. La estética	35
2.2. Percepción visual	38
2.3. Medición del valor estético	40
2.3.1. Medida del valor estético de Birkhoff	41
2.3.2. Medida del valor estético de Eysenck	42
2.3.3. Pruebas psicológicas de juicio estético	43
3. Estética computacional	45
3.1. Sistemas computacionales	47
4. Datasets	55
4.1. Datasets utilizados en Estética Computacional	56
4.1.1. PHOTO.NET (2006)	56
4.1.2. DPCHALLENGE.COM (2006)	57
4.1.3. MIR FLICKR (2008)	58
4.1.4. Dataset utilizado en Psicología (2009)	59
4.1.5. CUHKPQ (2011)	59

4.1.6. AVA (2012)	59
4.1.7. AADB (2016)	60
4.1.8. Thömmes y Hübner (2020)	60
4.1.9. IDEA (2020)	61
II Resultados	63
5. Dataset para la medición estética con Machine Learning	65
5.1. Limitaciones de los datasets estudiados	65
5.2. Propuesta de una nueva metodología	68
5.3. Construcción de un nuevo dataset	69
5.3.1. Datos originales	69
5.3.2. Dataset propuesto	71
5.3.3. Evaluación	72
5.4. Experimentación	75
5.4.1. Materiales y métodos	75
5.4.2. Modelos computacionales	76
5.4.3. Resultados	77
5.5. Discusión	81
5.6. Conclusiones	84
6. Comparación de modelos computacionales para la predicción estética	85
6.1. Introducción	85
6.2. Trabajos relacionados	86
6.3. Algoritmo híbrido de aprendizaje automático	88
6.3.1. Características de aprendizaje por transferencia	88
6.3.2. Método híbrido supervisado de aprendizaje automático (CGS)	89
6.4. Resultados	92
6.4.1. Comparativa con los métodos del estado del arte	94
6.5. Conclusiones	97
7. Validación de las propuestas en una aplicación práctica	99
7.1. Contexto	99
7.2. Trabajos relacionados	101
7.3. Aplicación comercial	102

7.4. Materiales y métodos	102
7.4.1. Dataset de entrenamiento	102
7.4.2. Dataset de evaluación	104
7.4.3. Muestra y procedimiento	107
7.5. Resultados	111
7.5.1. Análisis estadístico	114
7.6. Discusión	116
7.7. Conclusiones	119
III Conclusiones	121
8. Conclusiones y trabajo futuro	123
8.1. Conclusiones	123
8.2. Trabajo futuro	125
9. Conclusions and future work	127
9.1. Conclusions	127
9.2. Future work	129
Referencias	131
IV Apéndices	149
A. Producción científica	151
B. Galería de imágenes del dataset propuesto en el capítulo 5	163
C. Galería de imágenes utilizadas en el capítulo 7	177

Índice de figuras

4.1. Distribución de las puntuaciones de los datasets IDEA, AVA Y AADB. Imagen extraída de Jin et al. [80].	61
5.1. Resultados del sistema entrenado con el conjunto DPCv y validado con los diferentes subconjuntos descritos en la tabla 5.1.	66
5.2. Estructura de la metodología diseñada para la construcción de un nuevo dataset.	68
5.3. Caracterización de las 44.047 imágenes obtenidas inicialmente de DPChallenge. (a) disposición del número de votos dentro del rango de valoraciones válidas y (b) distribución de las evaluaciones medias de las imágenes dentro del rango de valoraciones válidas.	71
5.4. Caracterización de las 1.000 imágenes del conjunto propuesto. (a) Distribución del número de votos dentro del rango de puntuación y (b) distribución de las valoraciones medias en las imágenes dentro del rango de puntuación válido.	71
5.5. Distribución de las valoraciones medias de (a) estética y (b) calidad obtenidas en el grupo de control.	74
5.6. Gráficos de dispersión entre los rangos para las tres combinaciones posibles dados los criterios evaluados para todo el estudio.	74
5.7. Resultados obtenidos para los cuatro modelos propuestos y optimizados mediante hiperparametrización. A la derecha, se muestran los valores medios de Spearman (arriba) y Pearson (abajo). A la izquierda, las distribuciones de las 50 ejecuciones independientes de cada modelo óptimo (Spearman arriba y Pearson abajo) con diferentes tamaños de entrada probados mediante FS.	79

5.8. Distribución de las correlaciones obtenidas para cada modelo optimizado (Pearson a la derecha y Spearman a la izquierda). Para cada par, se muestra el valor p obtenido mediante una prueba de Kruskal-Wallis.	80
5.9. Distribución de las correlaciones (de Spearman a la derecha y de Pearson a la izquierda) obtenidas para cada uno de los tres criterios (DPChallenge, Estética y Calidad) utilizando 256 variables de entrada y un modelo SVM optimizado mediante hiperparametrización.	80
5.10. Ejemplos de imágenes con diferentes puntuaciones basadas en los tres criterios de evaluación. Para cada imagen, se da un valor numérico según cada criterio, mientras que las barras muestran el peso normalizado de dicho valor dentro de cada rango de evaluación (DPChallenge en el rango [1, 10] y estética y calidad en el rango [1, 5]).	83
6.1. Las 12 transformaciones matemáticas disponibles que se utilizaron para determinar el proceso evolutivo.	90
6.2. Diagrama de flujo de trabajo del algoritmo genético híbrido.	91
6.3. Resultados experimentales de ambos modelos híbridos de aprendizaje automático CNN-CGS comparados con el estado del arte.	93
6.4. Resultados de 50 ejecuciones independientes en AUROC para GLMNET, Random Forest y Support Vector Machines con GoogleNet (verde) y ResNet (violeta)	95
7.1. Diagrama que muestra el proceso de creación del dataset de evaluación.	105
7.2. Imágenes de los nueve primeros anuncios del conjunto original de una de la Agencia1.	106
7.3. Imágenes del conjunto estético de la Agencia1 propuestas para las primeras nueve posiciones.	107
7.4. Ejemplo de la encuesta desde la perspectiva de los participantes.	109
7.5. Imágenes con más votos afirmativos del conjunto original de la Agencia1.	113
7.6. Imágenes con más votos afirmativos del conjunto estético de la Agencia1.	114

- 7.7. Porcentaje de imágenes de cada conjunto por cada número de clics. El eje x define el número de clics con un máximo de 50 y el eje y representa el porcentaje de imágenes de cada conjunto que recibieron esos clics. Verde claro: imágenes de los conjuntos originales. Verde oscuro: imágenes de los conjuntos estéticos. 116
- 7.8. Boxplot en el que se muestra la distribución de ambos conjuntos de datos. Verde: clics del conjunto original. Violeta: clics del conjunto estético. 117

Índice de cuadros

3.1. Resumen de los trabajo estudiados.	52
5.1. Descripción de los conjuntos de datos utilizados para estudiar su grado de generalización en Carballal et al. [20].	66
5.2. Datos descriptivos del conjunto de datos original.	70
5.3. Datos descriptivos de cada uno de los cinco conjuntos de 200 imágenes que componen el dataset propuesto.	72
5.4. Media de los resultados presentados en la figura 5.7, identificando los hiperparámetros y el tamaño de la entrada para cada modelo.	78
6.1. Métodos y resultados del estado del arte en clasificación y predicción de estética usando el dataset de Photo.net [35] y empleando <i>accuracy</i> , AUROC y el valor de correlación Pearson como estándares de calidad.	88
6.2. Parámetros con los que se encontraron los mejores resultados de CGS con GoogleNet y ResNet.	93
6.3. Parámetros con los que se encontraron los mejores resultados para los métodos del estado del arte utilizando GoogleNet y ResNet.	96
7.1. Número de imágenes repetidas en el conjunto original y en el conjunto estético por cada caso de estudio.	108
7.2. Datos personales proporcionados por los participantes de la encuesta.	110
7.3. Número medio de votos recibidos por imagen para cada conjunto, el incremento por cada portal y el incremento medio total en los tres portales. El número máximo de votos por imagen es 50.	111

- 7.4. Número medio de votos recibidos por imagen para cada conjunto, el incremento para cada agencia y el incremento medio total en las tres agencias inmobiliarias. El número máximo de votos por imagen es 50. 112
- 7.5. Datos estadísticos para cada conjunto: el conjunto de imágenes original y el estético. 115
- 7.6. Media de votos por imagen recibido por cada conjunto de las agencias según el porcentaje de imágenes analizadas. 119

Abreviaturas

- **AADB**: Aesthetics and Attributes Database
- **ACQUINE**: Aesthetic Quality Inference Engine
- **Ada-Boost**: Adaptative Boosting
- **AMT**: Amazon Mechanical Turk
- **ANN**: Artificial Neural Networks
- **AUROC**: Area Under the Receiver Operating Characteristic
- **AVA**: A Large-Scale Database for Aesthetic Visual Analysis
- **CART**: Classification and Regression Trees
- **CGS**: Correlation by Genetic Search
- **CNN**: Convolutional Neural Network
- **CV**: Cross-Validation
- **DJT**: Design Judgment Test
- **DNN**: Deep Neural Network
- **FS**: Feature Selection
- **GBM**: Generalized Boosted Models
- **GLM**: Generalized Linear Model
- **GMM**: Gaussian Mixture Model
- **GSP**: Gaze Shifting Paths

- **IA:** Inteligencia Artificial
- **ILSVRC:** ImageNet Large Scale Visual Recognition Challenge
- **IS-MCA:** International Society for Mathematical and Computational Aesthetics
- **k-NN:** k-Nearest Neighbors
- **K-S:** Kolmogorov-Smirnov
- **ML:** Machine Learning
- **MRSSE:** Mean Residual Sum of Squares Error
- **NMF:** Factorización de matrices no negativas
- **RLM:** Regresión Lineal Múltiple
- **RMSE:** Root Mean Squared Error
- **RF:** Random Forest
- **SGD:** Descenso de Gradiente Estocástico
- **SIFT:** Scale-Invariant Feature Transform
- **SVM:** Support Vector Machine
- **SVM-RBF:** Support Vector Machine Radial Basis Function
- **SVM-RFE:** Support Vector Machine Recursive Feature Elimination
- **SVR:** Support Vector Regression
- **WND-CHARM:** Weighted Neighbor Distance using Compound Hierarchy of Algorithms Representing Morphology



Parte I

Estado del Arte

Capítulo 2

Estética y percepción

El concepto de estética y el valor estético de imágenes digitales son uno de los fundamentos sobre los que se construye esta tesis doctoral. Así, resulta esencial introducirnos en este campo a partir de su origen e historia, comprendiendo la percepción humana de este concepto y estudiando también las investigaciones psicológicas más relevantes que tratan de medir el valor estético de una imagen.

La estética del arte es históricamente un tema de disputa entre filósofos y teóricos del arte. En este capítulo nos adentraremos también en este campo para entender los principios estéticos de un arte como el de la fotografía.

2.1. La estética

El término estética deriva de la palabra griega *aisthesis*, que significa sensación o percepción, y en su origen se refería a las impresiones sensoriales. Este concepto aparece muy pronto en la historia de la filosofía, pues Platón ya planteó en su obra *Hippias Mayor* el problema de lo *bello*.

En 1753, el término *estética* adquirió un nuevo significado cuando en Alemania, Alexánder Gottlieb Baumgarten (1714-1762) identificó el conocimiento sensible con el conocimiento de la belleza, y denominó el estudio del conocimiento de la belleza con el nombre de estética [156]. A partir de este momento la estética afecta a las artes, pero no se restringe solo a ellas, pues se puede argumentar que la mayor parte de los elementos que forman parte de la vida cotidiana tienen su valor estético.

David Hume, filósofo francés, habló de *gusto* en lugar de estética, y Freeland [52] lo definió más adelante como “la refinada capacidad de percibir la calidad

en una obra de arte”. Hume, en su *Teoría del gusto*, destaca ciertos factores que pueden ayudar a encontrar la *norma del gusto*, elementos que permitirán contar con criterios objetivos para identificar la belleza. Hume afirma que para desarrollar el gusto será necesaria la práctica de un arte en particular, la comparación entre diferentes clases de belleza, manteniendo la mente libre de todo prejuicio.

El gusto es el objeto de la estética. El gusto se puede formar y se puede afinar, puede recibir una calificación positiva o negativa y ofrece sus juicios con un objetivo de universalidad, independientemente de la subjetividad. El gusto es, por tanto, una forma de conocimiento que no debe identificarse ni con el conocimiento sensible ni con el intelectual. El gusto es la capacidad de percibir la belleza, un sentimiento de placer ante algo bello que se presenta. Se entiende por gusto tanto el sentimiento en sí mismo como la facultad de percibirlo.

En la estética contemporánea, tanto la valoración de las obras de arte como la validez de los juicios sobre ellas no se establecen de acuerdo a unos principios objetivos, sino subjetivos. Sin embargo, para Hume, no por ser subjetivos carecen de validez universal. Se basa en la igualdad fisiológica y psicológica del género humano para fundamentar la posibilidad de que exista un gusto con carácter general, es decir, que pueda ser comprendido por cualquier persona, aunque no necesariamente aceptado. Hume hacía hincapié en la educación y en la experiencia: las *personas de gusto* adquieren capacidades que conducen a un acuerdo sobre qué autores y obras de arte son mejores. Cuando esas personas llegan a un consenso, se establece un *criterio del gusto* que es universal. Actualmente, existen ideas contrarias a este enfoque que alegan que estos *árbitros del gusto* de Hume solo adquirirían sus valores mediante adoctrinamiento cultural.

Se puede considerar adoctrinamiento cultural o se puede considerar aprendizaje, pero no se puede obviar que el nivel educativo influye en la preferencia estética. Existe una correlación inversa entre el nivel educativo y la tendencia a la elección de imágenes representativas de la realidad, lo que puede concluir en que la estética es, al menos en parte, aprendida [50].

Más tarde, Kant, otra de las grandes figuras de la filosofía, también habló de juicios de gusto [52], sin embargo, estaba más interesado en explicar los juicios sobre la belleza. Pretendía demostrar que los juicios en estética estaban basados en características de las propias obras, no solo en el ser humano y sus preferencias. Kant trató de describir las capacidades humanas para percibir y clasificar el mundo que nos rodea. Observó que, por norma general, el ser humano aplica

etiquetas o conceptos al mundo que le rodea para clasificar las entradas sensoriales que se ajustan a una finalidad. Los objetos bellos no sirven a finalidades humanas. Una flor bella agrada, pero no porque la queramos comer o porque necesariamente la queramos para formar un ramo. La forma en la que Kant reconoció esto fue decir que lo bello tiene *intencionalidad sin intención*. La flor tiene sin duda su propia finalidad (reproducir nuevas flores, por ejemplo), pero esa no es la razón de que sea bella. La disposición de sus características, como sus colores y texturas, impulsa a la mente humana a sentir que el objeto es *como debe ser*. Definimos algo como bello porque promueve una armonía interna en nuestra mente y suscita placer. Aunque esa definición es impulsada por un sentimiento o consciencia personal y subjetiva del placer, tiene supuestamente una aplicación objetiva. Kant explicó también que, para apreciar la belleza, la respuesta debe ser desinteresada, independientemente de su propósito o de las sensaciones placenteras que produzca.

En resumen, según Kant la estética se experimenta cuando algo estimula las emociones, intelecto e imaginación. Estas facultades se activan en un juego libre y no de forma deliberada. Lo bello atrae a nuestros sentidos, pero de forma fría y distante. La forma y el diseño de un objeto bello son la clave de la *intencionalidad sin intención*.

Korsmeyer, en su libro *Aesthetics: The Big Questions* [91], resume los estudios filosóficos de la estética en dos doctrinas opuestas, una que conceptualiza la estética como una cuestión de placer y gusto, y otra que considera la estética una cuestión de cognición y, por tanto, de procesos de aprendizaje. Korsmeyes explica la primera refiriéndose a filósofos como Platón y Kant y, la segunda, refiriéndose a Aristóteles y a otros filósofos contemporáneos como Hans-Georg Gadamer. Mientras que la primera doctrina ha sido la creencia principal hasta el siglo XVIII, la segunda está dominando los estudios más recientes.

A pesar de los numerosos estudios y enfoques que existen sobre el concepto de estética dentro de la filosofía y del arte, explicar qué es lo que hace que consideremos algo como estéticamente interesante, sigue siendo una tarea de gran dificultad. Dentro del ámbito artístico, se dice que la belleza estética se basa en la existencia de lo que se conoce como *principios estéticos*. Estos principios son básicos para la creación y posterior reconocimiento de cualquier tipo de obra de arte, ya que existe una estrecha conexión entre valor estético y percepción visual [55, 81, 105, 123, 144].

Dentro de estos principios estéticos destacan la armonía, la perspectiva, el ritmo y el movimiento. La armonía está relacionada con la unidad de una obra artística. En las artes, este principio implica simetría, proporción y equilibrio. La perspectiva aparece en nuestra vida cotidiana al intentar representar uno o varios objetos sobre una superficie plana. Con este principio se debe tener en cuenta el fenómeno visual con el que se perciben los objetos en realidad. Ritmo se refiere a la repetición periódica de una figura, que puede ser un movimiento virtual o real con distintas pausas o intervalos, definiendo así si se trata de un ritmo estático o dinámico. El movimiento pretende expresar la interacción de determinados objetos y cómo se complementan entre sí.

Otros conceptos visuales que se deben valorar en la apreciación estética son: (i) el centro de interés, que define la parte de la obra donde el autor pretende captar la atención del espectador; (ii) el espacio o la relación que existe entre los elementos utilizando la profundidad, el volumen y la superficie bidimensional; y (iii) la diferenciación entre figura y fondo.

2.2. Percepción visual

Como ya se adelantaba en la sección anterior, según Machado [105], existe una conexión muy estrecha entre valor estético y percepción.

El ser humano adquiere conciencia de sí mismo y del mundo que lo rodea por medio de sus sentidos. A partir de los estímulos recogidos por los sentidos las personas descubren, organizan y recrean la realidad, adquiriendo conciencia de ella por medio de la percepción.

El estímulo pertenece al mundo exterior y produce un primer efecto o sensación en la cadena del conocimiento. La percepción pertenece al mundo individual interior, al proceso psicológico de la interpretación y al conocimiento de las cosas y los hechos. La percepción es una interpretación significativa de las sensaciones.

La diferencia entre las sensaciones recibidas y la realidad del mundo físico que nos rodea la explica la psicología, aunque están implicadas también otras ciencias, como la geometría, la física o la biología.

Si nos centramos en un arte visual como la pintura o la fotografía, el análisis de la visión normalmente se realiza sobre un plano, donde se representan las formas de un mundo material o imaginado. Este fenómeno, que presenta un continente menor que el contenido, es un permanente desafío que abordan los artistas

y fotografías de todos los tiempos. Se mueven constantemente en el dilema de representar las cosas *como son* o *como se ven*, entre la representación de estímulos y sensaciones, y los resultados obtenidos por la percepción.

Aparte del conocimiento técnico que la representación requiere, cada concepción estética toma sus preferencias en este problema [33], y son las escuelas realistas, paradójicamente, las que se alejan más de la realidad perceptiva en la representación y se acercan al mundo de los estímulos, para que sus obras resulten reales en su percepción posterior. Como señala Cordero [33], deben representarse las cosas *como no son* para que parezcan *como son*.

En la percepción visual de las formas se produce un acto óptico-físico que funciona mecánicamente y de forma similar en todas las personas. Se trata de un proceso constructivo que empieza con la información sensorial, pero cuyo propósito no es representar el mundo tal como es, sino proporcionar una imagen que pueda ser comprendida y que sea coherente con nuestro previo conocimiento del mundo. Para esto, la percepción se guía por inferencias, hipótesis y otros procesos de arriba hacia abajo, como el contexto, que puede influenciar la apariencia de un objeto. Los psicólogos Gestalt caracterizan varios procesos perceptivos por medio de los que las características visuales se unen, se dividen y se agrupan para construir imágenes significativas [116].

Las diferencias comienzan con esta interpretación de la información recibida. Las desigualdades de cultura, educación, edad, memoria, inteligencia y hasta el estado emocional, pueden alterar en gran medida el resultado. Existen abundantes pruebas que indican que la captación de las formas y de los colores varía de acuerdo a la especie, el grupo cultural y el grado de entrenamiento del observador [3]. Así, lo que es racional para un grupo, será irracional para otro; es decir, no puede captarse, comprenderse, compararse o recordarse. Dentro de ciertos límites, el entrenamiento determina las categorías que son accesibles para cada individuo [3].

El acto perceptivo, aunque es algo cotidiano y se realiza de forma automática, no es nada simple y tiene múltiples implicaciones, ya que es evidente que el mundo real no es el que percibimos por la visión, y por esto es necesaria una interpretación constante y convincente de las señales recibidas. La percepción visual requiere de un aprendizaje que se va realizando a lo largo de toda la vida, aunque casi siempre de forma casual e inconsciente, por lo que sufre grandes alteraciones y condicionantes del medio en el que se ejercita.

Desde los tiempos de la Grecia clásica, la civilización occidental hizo hincapié en la distinción entre el mundo objetivamente existente y la percepción que de él se tiene. Esta distinción estableció la diferencia entre lo físico y lo mental y fue el comienzo de la psicología. Los pensadores griegos dieron por supuesto que el criterio para evaluar la percepción, provenía de la razón. La percepción sensorial y el razonamiento quedaban establecidos como antagonistas, que se necesitaban mutuamente [3].

Aunque los filósofos griegos concibieron la dicotomía de percepción y razonamiento, no se puede decir que aplicaran esta noción con la rigidez que la doctrina adquirió en los siglos recientes del pensamiento occidental. Los griegos aprendieron a desconfiar de los sentidos, pero nunca olvidaron que la visión directa es la fuente primera y última de la sabiduría. Refinaron las técnicas del razonamiento, pero también creyeron que, en palabras de Aristóteles, “el alma jamás piensa sin una imagen” [3].

2.3. Medición del valor estético

Durante varios siglos, las cuestiones relacionadas con el arte y la estética fueron respondidas mediante teorías y experimentos. La mayor parte de estos casos estaban basados en la observación de las reacciones de una población a determinados estímulos.

A finales del siglo XIX, algunos pensadores intentaron aplicar los métodos psicológico y experimental a la estética, pero resultó una tarea difícil por la fuerte relación de la estética con el sentimiento. Medir el sentimiento, el placer o la pena, es más difícil que medir fenómenos físicos. Fechner, considerado por muchos el fundador del campo de la psicofísica, fue el primero que trató de aplicar los métodos experimentales en estética. También fue el primero que realizó un estudio basado en la observación de las respuestas de sujetos representativos de diversas poblaciones a material visual diverso. Dichos experimentos sentaron las bases y los métodos experimentales para la formulación hipotética y su comprobación en condiciones controladas.

En su libro de introducción a la estética [47], Fechner enumera algunas leyes generales y principios de experiencia artística. Entre ellos, cabe destacar el *umbral estético*, por encima del cual solo lo agradable o desagradable se puede llamar así. A estos principios Fechner añade la afirmación de que siempre, y especialmente

cuando se trata de arte, nuestro encuentro con los signos externos (imágenes, palabras o sonidos) da lugar a un reconocimiento, porque desde siempre los cubrimos de recuerdos e impresiones personales, es decir, que están ligados a una experiencia anterior. Además, Fechner [47] destacó la importancia de dos fuerzas opuestas, la complejidad y el orden, y propuso que los estímulos son placenteros cuando tienen un balance adecuado de ambas fuerzas.

La aproximación experimental de Fechner a la estética inició una línea de investigación empírica sobre la conducta estética [126]. Elaboró estadísticas sobre las preferencias individuales hacia contornos regulares y otras formas simples, obteniendo unas curvas de distribución objetivas. Comparte con Taine [152] el mérito de haber introducido la estética *desde abajo*, es decir, basada en la observación sensorial, en oposición a la estética *desde arriba*, derivada de hipótesis metafísicas.

Por su parte, Daniel Berlyne desarrolló el primer programa de investigación enfocado al estudio de la estética, con el nombre *Estética y Psicobiología* [9]. Esta investigación destacó por la búsqueda de los elementos determinantes del juicio estético. Su objetivo era poder explicar mediante leyes hedónicas las preferencias estéticas de los seres humanos o animales ante determinados estímulos.

En 1916 nace una nueva corriente que propone representaciones icónicas y obras de arte pictórica como objeto de estimulación estética [44, 50, 88, 89]. Lo que interesa en estos momentos a los psicólogos es el problema del gusto.

Una vez presentada la base experimental de Fechner, era necesario determinar un método capaz de cuantificar la estética de un objeto. Determinadas características de las obras de arte, como la forma y situación de las líneas, las secuencias rítmicas o las variaciones de tono, son a partir de este momento objeto de medición. Primero Birkhoff y más tarde Eysenck propondrían las primeras fórmulas, que estudiaremos a continuación.

2.3.1. Medida del valor estético de Birkhoff

En 1932, Birkhoff propone la primera fórmula matemática que permitirá establecer una medida estética (M) para objetos visuales en función de su orden (O) y complejidad (C).

$$M = \frac{O}{C} \tag{2.1}$$

La ecuación 2.1 propone que cuanto más ordenados y sencillos sean los estímulos visuales asociados a un objeto, mayor será su belleza o valor estético.

Birkhoff [12] definió la complejidad como una expresión de multiplicidad, como el número de elementos que componen una imagen, mientras que el orden analiza la regularidad de esos elementos (repetición y redundancia). Birkhoff también expuso diferentes ejemplos visuales, pero no llevó a cabo experimentos para validar su hipótesis. A pesar de ello, existen diferentes trabajos de investigación que trabajan sobre su teoría, aunque no todos obtuvieron las mismas conclusiones. Por un lado, Brighouse [18] y Meier [121] concluyeron que la teoría de Birkhoff está fundada, mientras que, por el contrario, Weber [168], Beebe-Center y Pratt [8], Davis [38] y Eysenck [45] no están de acuerdo con su hipótesis.

2.3.2. Medida del valor estético de Eysenck

El estudio más completo relacionado con la hipótesis de Birkhoff fue desarrollado por Eysenck [44, 45]. Anteriormente, el propio Eysenck había realizado experimentos relacionados con dicha teoría, exponiendo su disconformidad con la misma.

Con el fin de proporcionar una fórmula alternativa para la medición estética, realizó su propio experimento en un entorno controlado. Este experimento contó con un total de 1.100 participantes, entre los que se encontraban personas con estudios relacionados con el arte y otras sin ellos (artistas, estudiantes, profesores y psicólogos). A esta población se le mostraron diferentes series de polígonos, que debían ordenar según su preferencia. Los polígonos de la prueba formaban parte del material que había utilizado Birkhoff en su investigación [12].

A partir de la prueba realizada, Eysenck presentó una fórmula diferente a la de Birkhoff, aunque también basada en la manifestación del orden y la complejidad. Sin embargo, al contrario que Birkhoff, la relación con la complejidad es positiva, ya que ambas, orden y complejidad, contribuyen positivamente en la apreciación de la belleza. Según Eysenck, la fórmula que ofrecía la mejor predicción sobre la preferencia humana era la siguiente:

$$M = O \times C \tag{2.2}$$

En ambos casos, tanto Birkhoff como Eysenck no contaban con medidas exactas, y mucho menos computacionales, que permitieran cuantificar el orden o la

complejidad.

2.3.3. Pruebas psicológicas de juicio estético

Al margen de las dos mediciones comentadas anteriormente, existen también pruebas psicológicas relacionadas con el juicio estético que pueden servir para su validación, ya que son relativamente objetivas, fácilmente reproducibles y proporcionan resultados cuantificables [4, 28, 29, 68, 73, 129, 141, 149]. El principal problema que se encuentra en este tipo de validación es la falta de consenso de los investigadores sobre las pruebas que se realizan. Los conceptos que se trabajan en dichas pruebas suelen ser discutibles, ya que se basan en propuestas del autor o autora y pueden no ser aceptados universalmente. Además, aunque varios autores concuerden en los principios estéticos elegidos, la capacidad de evaluación de una prueba en particular puede ser, y a menudo es, cuestionable. La media de resultados obtenidos por seres humanos en estas pruebas también varía de un estudio a otro. Aunque esto se puede deber, al menos en parte, por la selección de las personas participantes y otros factores exógenos, lo que hace más difícil entender qué es lo que constituye una buena puntuación en una prueba en particular. Además, la definición de muchos términos básicos para las tareas de juicio, como *estética*, *preferencia* o *expresión*, aún están del todo claras como afirma Child (1964).

Dentro de este tipo de pruebas destaca el “Design Judgment Test” (DJT), presentado por Maitland Graves en 1948 [61]. Este test se basa en ciertas teorías de la creación artísticas y la apreciación [62] y permite estimar ciertos componentes de la aptitud para evaluar estructuras artísticas. Para esto, se trata de estimar el grado de reconocimiento y la reacción de un individuo a principios básicos de orden estético, como es el caso de: unidad, predominancia, variedad, equilibrio, continuidad, simetría, proporción y ritmo. El DJT está compuesto por 90 elementos que pueden estar agrupados en pares o tríos de diseños semejantes. Uno de los diseños del grupo obedece a los principios enumerados y los restantes quiebran uno o más de esos principios. La tarea de la persona que realiza el test es escoger, en cada uno de los 90 elementos, el diseño que considera con mayor valor estético. Cada diapositiva se muestra durante unos 45-60 segundos y todas las láminas son de color blanco, negro y verde.

En las pruebas realizadas por Graves, el grupo de estudiantes de arte alcanza una mayor puntuación en comparación con el grupo de estudiantes que no cur-

saban estudios artísticos. Graves concluye que “la capacidad de la prueba para diferenciar los grupos artísticos de los grupos no artísticos es inequívocamente clara” [61]. En 1971, Eysenck y Castle [46] obtuvieron resultados diferentes que muestran una diferencia menor entre los estudiantes de arte y los estudiantes que no cursaban estudios de arte (64,4 % frente al 60 %), así como respuestas diferentes entre hombres y mujeres. Eysenck señala “el clima general de la enseñanza del arte, que ahora tiende a enfatizar la simplicidad y la regularidad en un grado mayor que hace 25 años” como una posible causa de las diferencias observadas.

Capítulo 3

Estética computacional

La estética computacional es un subcampo de la Inteligencia Artificial (IA) que se ocupa de la evaluación computacional de la belleza en dominios de la expresión creativa humana como la música, el arte visual o la poesía [78]. También se define como el campo que investiga los métodos computacionales capaces de tomar decisiones estéticas aplicables de manera similar a la de los humanos [131]. Normalmente, las fórmulas matemáticas que representan características o principios estéticos se utilizan junto con algoritmos especializados y técnicas estadísticas para proporcionar evaluaciones estéticas numéricas. Lo ideal es que se demuestre que esas evaluaciones se correlacionan bien con la evaluación humana. La estética computacional también puede mejorar la comprensión de la percepción estética humana.

Tal como se explica en la sección 2.3 del capítulo anterior, con Birkhoff nació la posibilidad de cuantificar el valor estético, lo que dio pie a la aparición diferentes líneas de investigación. Uno de estos nuevos campos de investigación fue la estética computacional.

En los años 50, el filósofo alemán Max Bense y, de forma independiente, el ingeniero francés Abraham Moles combinaron el trabajo de Birkhoff con la teoría de la información del ingeniero estadounidense Claude Shannon para idear un medio científico que pudiera comprender la estética [78]. Las ideas de Bense, a las que llamó *estética de la información*, y las de Moles, influyeron en algunas de las primeras obras generadas por ordenador.

A principios de la década de 1990 se fundó la Sociedad Internacional de Estética Matemática y Computacional (IS-MCA), especializada en el diseño que cuidaba la funcionalidad y la estética, y que pretendía ser un puente entre la ciencia y

el arte. A principios del siglo XXI, la estética computacional se había consolidado lo suficiente como para contar con sus propias conferencias especializadas, talleres y números especiales de revistas. La estética computacional sigue atrayendo, a día de hoy, a investigadores de diversos campos como la Inteligencia Artificial, la psicología, el arte o el diseño.

La razón por la que la evaluación estética de imágenes sigue siendo un importante campo de investigación es por su potencial y su riqueza en aplicaciones prácticas. El principal problema suele ser que la evaluación siempre va unida a la identificación del contenido semántico de las imágenes [125], lo que hace que la evaluación de la calidad estética sea una tarea difícil en términos de aislamiento de su contenido.

Obviamente, esto hace que la evaluación de la calidad estética de las fotografías sea también un reto para modelos informáticos. En los últimos años, muchos investigadores de diferentes campos del conocimiento se han centrado en la identificación de las características que más se correlacionan con las preferencias humanas, así como en la creación de sistemas informáticos capaces de recrear evaluaciones humanas para tareas de clasificación y predicción. Para resolver este problema, el estado del arte emplea conjuntos de datos creados a partir de portales fotográficos. El investigador obtiene un conjunto de imágenes y sus votos del portal fotográfico. Esta información contiene el valor estético medio de cada imagen. Hay algunos trabajos que se centran en la predicción de este valor y otros que realizan una clasificación binaria. Estos últimos se centran en dividir un conjunto de imágenes en dos conjuntos disjuntos que representan imágenes de alto valor estético frente a otro de bajo valor estético.

Para resolver estas tareas, el investigador suele emplear una arquitectura con dos componentes: un sistema que extrae las métricas de las imágenes y un sistema de Machine Learning (ML). La extracción de características es uno de los primeros pasos en cualquier problema con imágenes. Es vital para un buen rendimiento, ya que es la única información de entrada que se proporciona a los sistemas de ML. Las métricas son creadas ad-hoc por los investigadores basándose en la suposición de que dichas métricas están vinculadas al valor estético.

Los algoritmos de aprendizaje automático deben encontrar y filtrar las características que mejor permitan distinguir entre diferentes clases, conservando al mismo tiempo el poder de los detalles invariables dentro de la misma clase. Anteriormente, se han utilizado algoritmos de aprendizaje automático muy diferentes

para la evaluación estética, entre ellos Support Vector Machines (SVM), redes de neuronas artificiales (ANN) o modelos de mezcla gaussiana (GMM).

3.1. Sistemas computacionales

El aprendizaje automático consiste en extraer el conocimiento de los datos [124]. Es un campo de investigación que combina la estadística, la inteligencia artificial y la informática.

Las aplicaciones de las técnicas de aprendizaje automático se han hecho más populares en los últimos años en la vida cotidiana, desde el sistema de recomendación automática hasta la radio online personalizada, pasando por el reconocimiento de una persona en una foto, el coche que se conduce solo, el análisis de secuencias de ADN, la predicción de terremotos, etc [39]. Los algoritmos de aprendizaje automático más exitosos se clasifican en tres categorías: aprendizaje supervisado, aprendizaje no supervisado y aprendizaje por refuerzo . A continuación, se mencionan las técnicas de aprendizaje automático que se han utilizado en los últimos trabajos de evaluación estética de fotografías:

- K-Nearest Neighbors Classifier – Gehler y Nowozin [58], Zhu y Moh [176]
- Support Vector Machine (SVM) – Bishop et al. [13], Datta et al. [35], Datta y Wang [37], Marchesotti et al. [114], Nishiyama et al. [133], Tang et al. [154], Tian et al. [159], Wang et al. [167], Wong y Low [170], Zhu y Moh [176]
- Naive Bayes Classifiers – Csurka et al. [34], Ke et al. [87], Li y Chen [96]
- Adaptive Boosting (Ada-Boost) – Su et al. [150], Tong et al. [161]
- Classification and Regression Trees (CART) – Ciesielski et al. [32]
- Random Forest – Ciesielski et al. [32]
- Rank Support Vector Machine – Yeh y Chuang [173]
- Support Vector Regression – Bhattacharya et al. [10], Yeh y Chuang [173]

Actualmente, debido al generoso uso de dispositivos de imagen digital y cámaras, existe una enorme cantidad de fotografías digitales almacenadas en la nube,

en sitios como Google Drive, Facebook, Instagram o en comunidades de intercambio de fotografías online como Photo.net, DPChallenge o Flickr. La evaluación estética de esta enorme cantidad de fotografías se convierte en un trabajo difícil y, para resolverlo, recientemente se utiliza una arquitectura innovadora: las redes de neuronas profundas (DNN) [65, 71, 79, 80, 84, 85, 95, 102, 109]. Esto mejora la precisión de la tarea de evaluación.

Como se puede observar en la lista anterior, SVM y DNN son los métodos más utilizados en los trabajos de clasificación y evaluación estética de los últimos años. A continuación, se estudian con más detenimiento algunos de los enfoques más destacados, ordenados cronológicamente.

En 1998, Machado y Cardoso [106] propusieron por primera vez la utilización de la compresión de la imagen para estimar la percepción humana de complejidad y el valor estético. En este trabajo recurren a la compresión JPEG y fractal para juzgar el valor estético de las imágenes. Para evaluar la propuesta, los autores someten su sistema al “Design Judgment Test” [61]. El porcentaje de respuestas correctas obtenidas por el sistema depende de su parametrización, y va desde el 54,4% al 73,3% con una media del 64,9% sobre el intervalo paramétrico considerado.

Tong et al. [161] proponen una aplicación para la evaluación de la calidad integral como una extensión de la clasificación de imágenes. Se centran en una tarea de clasificación de imágenes de acuerdo a si fueron tomadas por un fotógrafo profesional o por un usuario sin experiencia fotográfica. Crean un conjunto de características de bajo nivel y las comparan con otras características de bajo nivel de uso general. Como conjuntos de entrenamiento emplean 16.643 imágenes de Corel Photo Stock y Microsoft Office Online catalogadas como “imágenes profesionales”, y 12.897 imágenes obtenidas del personal de Microsoft Research Asia catalogadas como “fotografías amateur”. A partir de dichas imágenes obtienen los valores utilizados durante la fase experimental. Como modelos de clasificación utilizan Ada-Boost, Real-AdaBoost, SVM y un clasificador Bayesiano, realizando 5 ejecuciones independientes para cada uno de ellos. Sus tasas de error fueron de 8,9%, 6,6%, 6,1% y 4,9% para cada clasificador respectivamente.

Datta et al. [35] extraen algunas características visuales basadas en la intuición, capaces de distinguir las imágenes estéticamente agradables de las desagradables. Intentan explorar la relación entre las emociones que las imágenes despiertan en la gente y su contenido a bajo nivel. Crearon un conjunto de en-

trenamiento formado por 3.581 imágenes procedentes del portal Photo.net que se analizará con más detalle en el capítulo 4. Utilizando los valores obtenidos a partir de las métricas que proponen seleccionaron mediante un análisis de componentes principales las 15 métricas más representativas. Con dicho subconjunto de métricas alcanzaron un 70,12% de acierto en la clasificación global utilizando Support Vector Machines (SVM), siendo un 60,08% en el caso de las imágenes con valores estéticos altos y un 72,31% en el caso de las imágenes con valores estéticos bajos.

Ke et al. [87] diseñaron un conjunto de *high level semantic features* para medir las diferencias perceptuales de diferentes imágenes. Al contrario que en aproximaciones como las de Tong et al. [161], en las que se usan conjuntos de *low level features* mediante una caja negra, ellos detallan un framework de resolución para la clasificación de imágenes. El dataset utilizado estaba compuesto por imágenes de la web DPChallenge.com y también será analizado posteriormente en el capítulo 4. Ordenan este conjunto de imágenes según su valoración media y determinan distintos conjuntos de entrenamiento separando el 10%, 8%, 6%, 4% y 2% superior e inferior. Utilizan un clasificador Naive-Bayes y alcanzan una tasa de error del 28%, 26%, 24%, 23% y 19% para cada porcentaje de imágenes respectivamente.

Luo y Tang [104] se basan en aspectos clave a la hora de realizar una foto profesional, como la composición, la luz o el color. Realizan un preprocesado que denominan “Subject Region Extraction” a partir del cual identifican el primero plano (a lo que se refieren como “subject region”) y el fondo (“clear region” o “background”) de cada imagen mediante filtros de detección de bordes y crean para cada uno de ellos una nueva imagen. Desarrollan cinco nuevas métricas que aplican a estas dos nuevas imágenes. Utilizando dicha información realizan su modelo de clasificación sobre el conjunto de datos de entrenamiento creado por Ke et al. [87]. Con esto, alcanzaron un porcentaje de acierto superior al 80% con cualquiera de las métricas propuestas. En concreto, la métrica relativa al “clarity contrast” por sí sola produce más del 85% de acierto. Los autores señalaron que la diferencia entre sus resultados y los de Ke et al. [87] se puede deber a la aplicación de métricas a las regiones de fondo de la imagen y a la mayor adecuación de la propia métrica.

Por otra parte, Wong y Low [170] crean distintas métricas agrupadas en tres categorías de características globales basadas en (i) técnicas básicas; (ii) normas

fotográficas; y (iii) ajuste de cámara. También emplean métricas de alto nivel basadas en regiones salientes y las relaciones entre sujeto y fondo. En su caso, utilizan el dataset creado por Datta et al. [35]. La clasificación la realizan mediante SVM con un kernel lineal. Utilizan una validación cruzada mediante 5 ejecuciones independientes. Son capaces de clasificar correctamente el 78,2% de las imágenes, correspondiente con el 82,9% para aquellas consideradas de alta calidad estética frente al 75,6% de las de baja calidad estética.

Liu et al. [100] trataron de identificar un conjunto de reglas de composición e implementarlas computacionalmente para permitir una evaluación cuantitativa y, además, buscaron facilitar una herramienta de edición automática de imágenes que mejorase la estética de una fotografía y la experiencia fotográfica cotidiana del usuario. Para ello, se centraron en las propiedades estéticas de la composición de imágenes y emplearon reglas reconocidas en la comunidad fotográfica. Los autores desarrollaron y formalizaron la puntuación estética basándose en un conjunto de directrices primarias de composición, que incluyen la regla de los tercios, la dominancia diagonal, el equilibrio visual y el tamaño de la región. Este trabajo supuso el primer intento de incorporar las directrices de dominancia diagonal, equilibrio visual y tamaño de la región en una puntuación estética automática. Para modificar la composición de las fotografías utilizaron un operador compuesto de recorte y reajuste. El operador de recorte selecciona un subconjunto de objetos de la imagen, y luego el operador de reajuste ajusta sus ubicaciones relativas. Al buscar una combinación de parámetros que produzca la imagen con la máxima puntuación estética, generan una imagen de salida que es una versión mejorada de la original.

Datta y Wang [37] implementaron un sistema de evaluación de imágenes al que denominaron Aesthetic Quality Inference Engine (ACQUINE). Con ello, proponen una puntuación estética entre 0 y 100 a las imágenes subidas por el usuario. Siguiendo trabajos anteriores, el motor de ACQUINE se basa en dos clases del clasificador SVM entrenado con un conjunto de datos compuesto por más de 20.000 fotografías y valoraciones obtenidas de Photo.net.

Bhattacharya et al. [10] presentan una aplicación interactiva que permite a los usuarios mejorar la estética visual de sus fotografías digitales mediante la recomposición espacial. Para evaluar la calidad estética de las fotografías utilizaron un modelo Support Vector Regression (SVR) que entrenaron a partir de los datos de los usuarios y con el que obtuvieron una precisión del 86%.

Romero et al. [140] exploraron el uso de estimaciones de complejidad para predecir el valor estético de fotografías. Utilizaron un conjunto de métricas de estimación de la complejidad basadas en el tamaño de codificación y el error de compresión JPEG y Fractal, que se aplican al canal de valor original y a las imágenes resultantes de aplicar los filtros Sobel y Canny a este canal. Utilizaron dos clasificadores diferentes (SVM y ANN) entrenados con el dataset creado por Datta et al. [35]. Alcanzaron el mejor resultado de clasificación utilizando SVM con un valor de precisión del 74,59%.

Lu et al. [102] utilizan redes de neuronas convolucionales profundas (CNN) para resolver tres problemas diferentes: el reconocimiento del estilo de la imagen, la categorización de la calidad estética y la estimación de la calidad de la imagen. La red profunda que proponen integra el aprendizaje de características compartidas y el aprendizaje de la función de agregación en un único marco unificado. Para el problema de la categorización estética utilizaron el dataset AVA y siguieron la misma configuración experimental que Murray et al. [127]: utilizaron 230.000 imágenes para el entrenamiento y 20.000 para test. Las imágenes de entrenamiento se dividieron en dos categorías: imágenes de bajo valor estético e imágenes de alto valor estético, basándose en los mismos criterios también que Murray et al. [127]. Los resultados de Lu et al. [102] concluyen que la red que incluye múltiples parches de la imagen en diferentes columnas genera un mejor rendimiento de predicción que el entrenamiento de una red con un solo parche de la imagen, alcanzando una precisión del 75,4%.

Mai et al. [109] utilizan, igual que Lu et al. [102] CNN y el dataset AVA en su propuesta experimental, pero con un enfoque diferente. El objetivo de Mai et al. [109] es poder preservar la composición de la imagen en la entrada de las CNN, en lugar de que la imagen tenga que ser transformada mediante recorte, escalado o relleno para adaptarse a un tamaño fijo de entrada. Para ello, presentan un nuevo método que añade una capa de *pooling* espacial adaptativo sobre las capas de convolución para manejar directamente las imágenes de entrada con tamaños y relaciones de aspecto originales. Los autores utilizaron la red VGG preentrenada con el conjunto de datos ImageNet y su método basado en CNN para la evaluación estética de imágenes y obtuvieron una precisión del 77,1%.

Kao et al. [84] abordan la cuestión de la correlación entre la evaluación automática de la calidad estética y el reconocimiento semántico. Se basan en redes de neuronas convolucionales (CNN) y utilizan un marco multitarea único para

explorar la correlación potencial entre estética y semántica. Posteriormente, los autores sustituyen las seis primeras capas de su red multitarea por AlexNet, VGG Net y ResNet. Así, las redes se inicializan con modelos preentrenados en ImageNet y se afinan con los datos de entrenamiento etiquetados con etiquetas estéticas y semánticas. Utilizan el dataset AVA para el entrenamiento y alcanzan el mejor porcentaje de precisión utilizando ResNet (79,08 %).

Jin et al. [80] proponen un nuevo dataset para la evaluación estética de imágenes con una distribución uniforme, en contraposición a los existentes con una distribución fuertemente normal que puede provocar un sobreajuste en el entrenamiento. Para testar su conjunto de datos utilizaron una nueva arquitectura de red neuronal con conciencia de agregación espacial. Aplicaron un módulo *Squeeze-and-Excitation* a la red ResNet y consiguieron los mejores resultados entrenando esta red con su nuevo dataset, con un valor MRSSE = 0,2856.

Cuadro 3.1: Resumen de los trabajo estudiados.

Trabajo	Dataset	Métodos utilizados	Objetivo principal	Resultados
Tong et al. [161] (2004)	Propio	Ada-Boost, SVM y clasificador bayesiano	Clasificación entre fotografía profesional y fotografía amateur	Error = 4,9 % (clasificador bayesiano)
Datta et al. [35] (2006)	Photo.net	SVM	Explorar la relación entre las emociones que suscita una imagen y su contenido de bajo nivel	Precisión = 70,12 %
Ke et al. [87] (2006)	DPChallenge	Clasificador Naive-Bayes	Clasificación de imágenes a partir de características de alto nivel	Menor error alcanzado = 19 %
Luo y Tang [104] (2008)	DPChallenge	Algoritmo propio	Clasificación de imágenes a partir de la identificación sujeto-fondo	Precisión = 85 %

Wong y Low [170] (2009)	Photo.net	SVM	Clasificación de imágenes a partir de características de alto y bajo nivel	Precisión = 78,2%
Liu et al. [100] (2010)			Creación de una herramienta que mejora la calidad estética de fotografías a partir de recorte y reajuste	
Datta y Wang [37] (2010)	Photo.net	Dos clases diferentes de SVM	Creación de un sistema de clasificación de imágenes (ACQUINE)	
Bhattacharya et al. [10] (2010)	Propio (Flickr)	SVR	Aplicación interactiva que permite mejorar la estética de fotografías mediante recomposición espacial	Precisión = 86%
Romero et al. [140]	Photo.net	SVM y ANN	Clasificación de imágenes a partir de métricas de estimación de la complejidad	Precisión = 74,59%
Lu et al. [102] (2015)	AVA	CNN	Reconocimiento del estilo, categorización estética y estimación de la calidad en imágenes	Precisión = 75,4% (categorización estética)
Mai et al. [109] (2016)	AVA	CNN VGG	Preservar la composición de la imagen en la entrada de las CNN	Precisión = 77,1%
Kao et al. [84] (2017)	AVA	CNN VGG, ResNet y AlexNet	Comprobar la relación entre el valor estético y el valor semántico de una imagen	Precisión = 79,08% (con ResNet)
Jin et al. [80] (2020)	IDEA	CNN ResNet	Crear un nuevo dataset con una distribución uniforme	MRSSE = 0,2856

Capítulo 4

Datasets

Una de las primeras necesidades detectadas en el estudio del estado del arte fue la de analizar con detenimiento los datasets utilizados en trabajos previos de predicción estética.

Normalmente, los investigadores dedican mucho esfuerzo a la creación y optimización de algoritmos para la predicción de parámetros en numerosos campos de la ciencia. Sin embargo, a veces no se presta el tiempo y la atención necesaria a la creación o elección del conjunto de datos que servirá para el entrenamiento y validación del sistema. Un buen conjunto de datos es esencial para obtener unos resultados óptimos con un sistema de Machine Learning.

Un conjunto de datos ideal para el entrenamiento de modelos de Machine Learning debería cumplir los requisitos que se presentan a continuación.

En primer lugar, el conjunto de datos debe ser representativo. En el pasado, los investigadores solían crear sus propios conjuntos de datos ad-hoc, quizá incluso a partir de sus colecciones personales de imágenes, lo que obvia la representatividad, ya que resulta necesario contar con muchos puntos de muestra diferentes, que cubran en conjunto todo el espectro de las fuentes de imágenes. Lo ideal sería que miles de personas contribuyeran al conjunto de datos y que este fuera lo suficientemente amplio para ser representativo del conjunto.

En segundo lugar, el dataset debería aportar información estadística sobre las evaluaciones recibidas en cada imagen. Idealmente, también aportaría información sobre los evaluadores y el proceso de evaluación. Esto permitiría conocer, por ejemplo, la distribución de datos con la que se entrena el sistema.

En tercer lugar, el conjunto de datos debería ser fácilmente accesible y de libre distribución. No deberían exigirse formularios de derechos de autor, y cualquier

investigador debería poder distribuir legalmente el conjunto de pruebas.

En cuarto lugar, es importante tener al menos una prueba estandarizada asociada al conjunto de datos para poder realizar evaluaciones comparativas.

4.1. Datasets utilizados en Estética Computacional

Existen numerosos datasets de imágenes digitales utilizados para el estudio de la Estética Computacional y de la complejidad. A continuación se realiza un pequeño análisis de los más utilizados, ordenados de forma cronológica.

4.1.1. PHOTO.NET (2006)

Datta et al. [35] crean un nuevo conjunto de datos con 3.581 imágenes procedentes de Photo.net, un sitio web que reúne más de un millón de imágenes pertenecientes a más de 400.000 usuarios. Cada imagen recibe una calificación de dos criterios: estética y originalidad. Estos dos criterios se valoran en una escala de 1 a 7. La información y las calificaciones pertenecientes a cualquier imagen son de carácter público y pueden verse en el mismo sitio web.

Photo.net publica parte de la información de las imágenes, mostrando con cada una las calificaciones de originalidad y estética en una escala media y los valores de sus características. Sin embargo, estos datos no incluyen información sobre los evaluadores.

Cada imagen del total de las 3.581 fue valorada por al menos 2 individuos diferentes, con una puntuación media situada en el intervalo de [3,55-7], con una media global que asciende a 5,06 y una desviación estadística de la normal de 0,83.

Datta et al., e igualmente otros investigadores como Wong y Low [170] que han utilizado este conjunto muestral, realizan una distribución preliminar con el objetivo de obtener dos grupos diferentes, empleando las valoraciones estéticas de los usuarios: un grupo de imágenes de alta calidad denominado *High* y otro de baja calidad denominado *Low*. Aquellas imágenes que han obtenido una puntuación mayor o igual a 5,8 se clasifican como *High* y las menores o iguales a 4,2 se clasifican como *Low*. Finalmente, se reúnen los dos grupos, el alto de 832 imágenes y el bajo, de 760.

Posteriormente, Datta et al. [36] publicaron un nuevo trabajo en el que presentaban otros cuatro conjuntos de datos que podían utilizarse en tareas de clasificación estética. Entre ellos, se encontraba un nuevo conjunto con imágenes procedentes de Photo.net. Este nuevo conjunto de datos se compone de 14.839 imágenes con una media de 12 valoraciones por imagen y una desviación típica de 13.

Datta et al. continuaron aumentando el número de imágenes de su conjunto muestral, manteniendo la web de origen [37]. Esta última ampliación permite un análisis estadístico más completo, ya que proporciona datos específicos sobre las valoraciones de cada imagen, indicando el número de votos para cada nivel de la escala de valoración. En este caso, cuenta con 20.278 imágenes, todas ellas valoradas por al menos cuatro personas. El rango de valoraciones medias se sitúa entre [2,33-6,90], con una media de 5,15 y una desviación de la normal de 0,58.

4.1.2. DPCHALLENGE.COM (2006)

El conjunto de datos publicado por Ke et al. [87] es uno de los más utilizados en los experimentos de clasificación estética de los que, por los resultados obtenidos, destaca Luo y Tang [104]. Se creó recopilando las imágenes del sitio web de fotografía DPChallenge.com y posee un total de 60.000 fotografías de 40.000 fotógrafos diferentes. A diferencia de las colecciones de Datta et al. obtenidas de Photo.net, todas las imágenes han sido valoradas por al menos 100 personas. Las valoraciones varían en una escala de entre 1 y 10 (siendo la primera la puntuación más baja posible y la segunda, la más alta).

DPChallenge es un sitio web de concursos fotográficos. Contiene un conjunto diverso de fotografías de alta y baja calidad de muchos fotógrafos diferentes. Las fotografías se agrupan por *challenges* o retos y la comunidad vota cada imagen en una escala de 1 a 10. A diferencia del portal Photo.net, en DPChallenge no se especifica qué cualidad o característica de la imagen se vota, los usuarios simplemente ejercen su juicio según su propio criterio y según el reto en el que esté incluido la fotografía.

Con el fin de trabajar con conjuntos más controlados, Ke et al. [87] crearon una colección con dos partes: High y Low. Del total, extrajeron el 10% con la valoración media más alta y el 10% con la más baja, de forma que cada subgrupo estaba compuesto por 6.000 imágenes. Posteriormente, realizaron una división mediante una dicotomía aleatoria con cada uno de los dos subgrupos, obteniendo

finalmente cuatro grupos de 3.000 imágenes cada uno, dos de alta y dos de baja calidad. Cada uno de estos grupos se empleó con el mismo fin: dos servirán para desarrollar los sistemas creados mientras que los otros dos validarán la eficacia y capacidad de los anteriores.

4.1.3. MIR FLICKR (2008)

Huiskes y Lew [74] crearon una nueva colección formada por 25.000 imágenes que fueron descargadas del sitio de fotografía social Flickr.com bajo licencia Creative Commons. Las licencias de este tipo permiten el uso de la imagen siempre que se acredite al fotógrafo por la creación original. Normalmente, el uso de las imágenes se concede bajo restricciones adicionales, que pueden incluir: solo se permite el uso no comercial, no se permite crear obras derivadas, o la distribución de obras derivadas solo se permite bajo la condición “Share Alike” (es decir, con una licencia idéntica a la original). Ninguna de estas restricciones impide el uso de estas imágenes para fines de evaluación comparativa.

Las 25.000 imágenes del conjunto original pertenecen a un total de 9.862 usuarios y 5.566 de ellos están representados en el conjunto por una única imagen. El usuario con el mayor número de imágenes en el conjunto tiene una contribución de 41 imágenes. Esta colección de imágenes es representativa para el dominio de las fotografías originales y de alta calidad.

Uno de los principales atractivos de Flickr es que la plataforma ofrece a sus usuarios la posibilidad de buscar y compartir sus imágenes basándose en etiquetas. El número medio de etiquetas por imagen es de 8,94. En la colección hay 1.386 etiquetas que aparecen en al menos 20 imágenes.

Los autores definen, además, retos estandarizados para que los investigadores puedan comparar los resultados. En los dos primeros retos (Visual Concept/Topic Recognition y Tag Propagation), la tarea consiste en el reconocimiento de patrones de detección de conceptos visuales, donde la suposición es que las etiquetas de la imagen están altamente correlacionadas con los conceptos visuales de la imagen. En el tercer reto (Tag Suggestion), investigan una gama muy amplia de etiquetas y se centran en la construcción de un sistema de recomendación de etiquetas con un orden jerárquico. El objetivo es maximizar la correlación entre las etiquetas clasificadas por el sistema automático y las etiquetas de los usuarios de Flickr.

Este dataset cuenta con una extensión realizada en el año 2010 [75] que lo

eleva a un total de un millón de imágenes con las mismas características que el original.

4.1.4. Dataset utilizado en Psicología (2009)

Cela-Conde et al. [24, 25] crearon un dataset formado por un conjunto final estandarizado de 800 imágenes divididas en 5 categorías: abstractas artísticas (AA), abstractas no artísticas (AN), representativas artísticas (RA), representativas no artísticas (RN) y fotografías de escenas naturales y construcciones humanas (NHS).

Las imágenes fueron mostradas a un grupo de 240 participantes (112 hombres y 128 mujeres, con edad media de 22,03 años y una desviación estándar de 3,75), dividido aleatoriamente en subgrupos de 30 personas. El experimento se realizó en un entorno experimental controlado en el que se mostraba a los participantes cada imagen durante 5 segundos en los que debían valorar la complejidad visual de un subconjunto de estímulos en una escala Likert de 1 a 5 (siendo 1 la peor valoración posible y 5, la mejor). De este modo, cada imagen recibió un total de 30 valoraciones. El valor medio obtenido por cada subgrupo para cada estímulo fue el valor considerado para representar la percepción de complejidad de ese estímulo en el conjunto final. Los estímulos de este conjunto fueron utilizados por Cela-Conde et al. [24], Nadal et al. [128], Machado et al. [107] y Castro [23].

4.1.5. CUHKPQ (2011)

El dataset CUHKPQ, creado por Luo et al. [103], está formado por 17.613 imágenes obtenidas de distintas fuentes online y divididas en 7 categorías semánticas. Cada imagen fue etiquetada como de alta o baja calidad por 10 evaluadores. El conjunto inicial estaba formado por 32.097 imágenes, pero solo se utilizaron en el conjunto final aquellas en las que la etiqueta coincidía en al menos 8 de los 10 participantes. Por tanto, se puede decir que este dataset está formado con etiquetas binarias de imágenes con un consenso alto.

4.1.6. AVA (2012)

Murray et al. [127] crearon el dataset *A Large-Scale Database for Aesthetic Visual Analysis* (AVA), una colección de imágenes y metadatos obtenidos de la plataforma DPChallenge, de la misma forma que Ke et al. [87]. En este caso,

los autores asociaron cada imagen con la información correspondiente del reto al que pertenecen en DPChallenge. Esta información puede ser explotada en combinación con puntuaciones estéticas o etiquetas semánticas.

AVA está compuesto por 255.000 imágenes incluidas en un total de 963 retos. Cada imagen cuenta con tres tipos de anotaciones diferentes: anotaciones estéticas, anotaciones semánticas y anotaciones de estilo de fotografía. En relación a las anotaciones estéticas, cada imagen está asociada a una distribución de puntuaciones que corresponden a los votos individuales recibidos en DPChallenge. El número de votos por imagen oscila entre 78 y 549, con una media de 210 votos por imagen.

4.1.7. AADB (2016)

Kong et al. [90] crearon el *Aesthetics and Attributes Database*, un conjunto de 10.000 imágenes procedentes de la plataforma Flickr con licencia Creative Commons. Comenzaron con la descarga de un conjunto amplio y variado de imágenes, y realizaron un procesado manual mediante el que eliminaron las imágenes no fotográficas (como caricaturas, dibujos, pinturas, imágenes de anuncios, etc.). A continuación, 5 evaluadores diferentes puntuaron de forma independiente cada imagen con una valoración estética general y un conjunto fijo de once atributos significativos utilizando Amazon Mechanical Turk (AMT). Cada imagen obtuvo su puntuación de valor estético a partir de la media de las cinco evaluaciones. Los autores comprobaron que las calificaciones medias se ajustaban bien a una distribución gaussiana.

Para la evaluación, dividieron aleatoriamente el conjunto de datos en validación (500 imágenes), test (1.000 imágenes) y entrenamiento (el resto).

4.1.8. Thömmes y Hübner (2020)

Thömmes y Hübner [158] propusieron recientemente un dataset formado por un total de 15.073 imágenes divididas en tres categorías diferentes: arquitectura, danza y paisaje.

Las imágenes de este conjunto se obtuvieron de Instagram. Para la construcción del conjunto de imágenes, los autores eligieron cuentas ya asentadas en la plataforma que cumpliesen tres criterios: (i) debían ser cuentas de fotógrafos profesionales que compartieran principalmente contenido de calidad; (ii) las cuentas

deben crear un contenido homogéneo que se clasifique como arquitectura, retratos de bailarines o fotografía de paisajes; y (iii) debían contar con al menos 15.000 seguidores, ya que un gran número de seguidores garantiza que una gran cantidad de usuarios ve y, potencialmente, le gustan las imágenes. Finalmente, se recogieron imágenes de 9 cuentas de fotógrafos profesionales. Las imágenes tienen una media de 3.228,85 *likes*.

En el experimento de Thömmes se utilizaron solo 270 imágenes del conjunto total (30 imágenes de cada cuenta). Cada participante evaluaba 32 imágenes aleatorias. Contaron con un total de 80 participantes y obtuvieron una media de 9,5 evaluaciones por imagen. La encuesta online, que fue realizada a través de la plataforma AMT, tenía una duración de 20 minutos y se pagaba a 3€. Se excluyeron a 7 de los 80 participantes por baja fiabilidad (< 0.50).

4.1.9. IDEA (2020)

El objetivo de Jin et al. [80] con este trabajo era el de crear un conjunto de imágenes para la evaluación estética con una distribución uniforme, en contraposición a los existentes con una distribución fuertemente normal, que puede provocar un sobreajuste en el entrenamiento. En la figura 4.1, se puede observar la distribución del dataset IDEA y de otros del estado del arte.

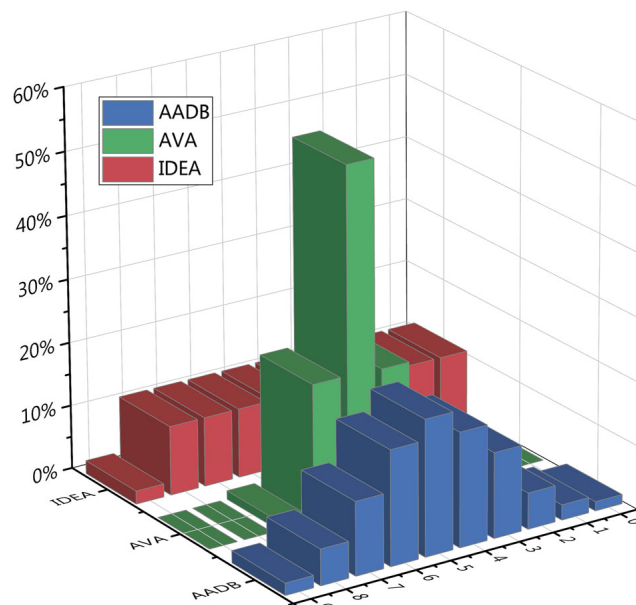


Figura 4.1: Distribución de las puntuaciones de los datasets IDEA, AVA Y AADB. Imagen extraída de Jin et al. [80].

Para crear este dataset, los autores recogieron imágenes y etiquetas de puntuación de los sitios web de fotografía DPChallenge y Flickr. Las puntuaciones de los sitios web van de 0 a 9, por lo que seleccionaron 1.000 imágenes para cada segmento (la etiqueta de puntuación de Flickr procede del dataset AADB).

Primero, trataron de seleccionar las 1.000 imágenes de cada rango en el portal DPChallenge. Eligieron las imágenes que tenían un mayor número de votantes en cada rango. A partir de este proceso, obtuvieron 7.961 imágenes. Faltaban, principalmente, imágenes con una puntuación de 1, de 2 y de 9 puntos, por lo que seleccionaron y complementaron aleatoriamente las imágenes con el conjunto de datos de AADB. Finalmente, no fue posible obtener 1.000 imágenes con una puntuación de 9, por lo que el conjunto final está formado por 9.191 imágenes.

Para el entrenamiento de modelos separaron aleatoriamente 1.000 imágenes para test y las 8.191 restantes se dedicaron a entrenamiento.

Parte II

Resultados

Capítulo 5

Dataset para la medición estética con Machine Learning

La primera parte experimental de esta tesis se basa en la propuesta de una nueva metodología para la construcción de nuevos conjuntos de imágenes digitales utilizados en clasificación y predicción de estética a través de Machine Learning. En primer lugar, se parte del análisis de diferentes conjuntos de datos frecuentemente utilizados para este propósito y explicados ya en el capítulo 4. Los resultados de dicho análisis concluyeron en la detección de diferentes problemas de composición en los conjuntos mencionados para trabajar con Machine Learning, por lo que se propone un nuevo conjunto de datos con origen en el portal fotográfico DPChallenge y, con ello, una nueva metodología extrapolable a otras fuentes de datos. Finalmente, se valida el dataset propuesto con diferentes sistemas de Machine Learning en términos de predicción estética.

5.1. Limitaciones de los datasets estudiados

Tras estudiar los datasets descritos en el estado del arte (capítulo 4), se llegó a la conclusión de que su capacidad de generalización no era suficiente para tareas de clasificación y predicción con Machine Learning. En muchos casos, el porcentaje de correlación es notablemente superior cuando el sistema es validado con el mismo conjunto de imágenes con el que ha sido entrenado. Sin embargo, en experimentos en los que el test se realiza con un dataset diferente al del entrenamiento, los resultados de correlación descienden notablemente. Esto fue comprobado en experimentos realizados anteriormente por el grupo de investigación que respalda

Cuadro 5.1: Descripción de los conjuntos de datos utilizados para estudiar su grado de generalización en Carballal et al. [20].

Origen	Publicación	Conjunto original	Conjunto disponible	Nombre
Photo.net	Datta et al. [35]	3.581	3.247	PN06
Photo.net	Datta et al. [36]	20.278	18.105	PN08
DPChallenge.com	Ke et al. [87]	6.000	6.000	DPCt
DPChallenge.com	Ke et al. [87]	6.000	6.000	DPCv

esta tesis doctoral [19, 20]: al entrenar el sistema con un subconjunto de 6.000 imágenes procedentes del dataset de DPChallenge creado por Ke et al. [87] el resultado de correlación es de un 91,38 % si se valida con el mismo subconjunto; sin embargo, este porcentaje desciende hasta un 60,26 % si se valida con un subconjunto diferente del mismo dataset, hasta un 56,21 % si se valida con el dataset creado por Datta et al. [35] en 2006 y hasta un 55,39 % si se hace con el creado por Datta et al. [36] en 2008. La descripción de los conjuntos utilizados en ese experimento y los resultados obtenidos se pueden observar en la tabla 5.1 y en la figura 5.1, respectivamente.

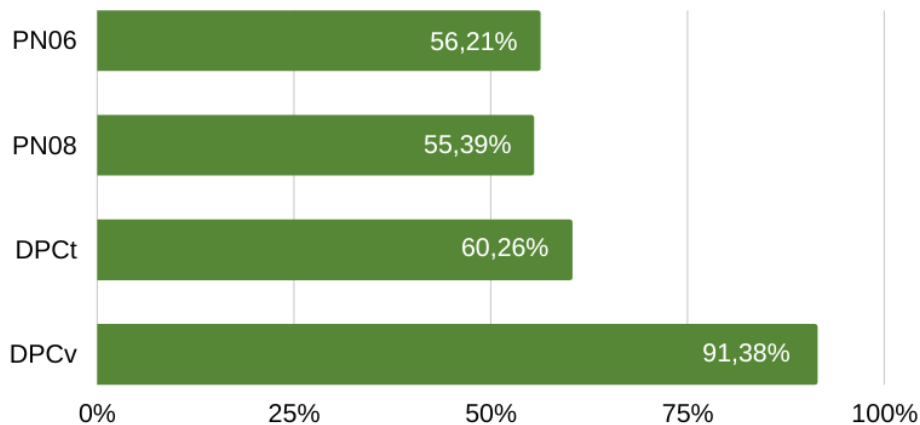


Figura 5.1: Resultados del sistema entrenado con el conjunto DPCv y validado con los diferentes subconjuntos descritos en la tabla 5.1.

Además de esto, los conjuntos de datos que cuentan con valoraciones procedentes de los portales fotográficos tienen otros defectos por los que pueden estar

sesgados:

1. El sistema de evaluación no tiene el mismo control que un test psicológico o una evaluación presencial. No es posible obtener toda la información sobre los evaluadores ni sobre el proceso de evaluación.
2. La cantidad de imágenes puede ser insuficiente para un problema de clasificación, pues no se justifica en ningún caso el tamaño de la muestra y existe una gran diferencia entre el número de personas que valoran cada imagen.
3. Las valoraciones de los usuarios pueden estar condicionadas por gustos personales o por la popularidad o auge momentáneo de ciertos estilos.
4. En uno de los casos [35], se ha demostrado que los usuarios de estos portales no tienen una base de conocimiento previa que les permita diferenciar entre criterios de estética y de originalidad.
5. Al tratarse la mayoría de datasets diseñados para clasificación binaria, solo se emplean aquellas imágenes evaluadas con puntuaciones extremas (las que reciben la puntuación más alta y las que reciben la puntuación más baja), dejando fuera del conjunto aquellas imágenes con evaluaciones intermedias

Por otra parte, en el dataset creado por Cela-Conde et al. [24] el número de imágenes presentado por categoría no es equitativo, por lo que no se pueden considerar los resultados obtenidos como representativos del conjunto. Además, parte de una cantidad importante de subconjuntos de imágenes, lo que produce que el dataset se convierta finalmente en varios datasets de menor tamaño y con menor coherencia interna.

Tras haber detectado estas limitaciones en los conjuntos existentes, se ha llevado a cabo la construcción de un nuevo dataset para la predicción del valor estético en imágenes, valorado con criterio humano en condiciones experimentales controladas a partir de un conjunto coherente de imágenes. En el proceso de construcción del dataset, se prestó especial atención al diseño de una metodología que se pudiera extrapolar a la construcción de otros conjuntos de datos y que pusiera solución a las limitaciones detectadas anteriormente.



Figura 5.2: Estructura de la metodología diseñada para la construcción de un nuevo dataset.

5.2. Propuesta de una nueva metodología

Con la intención de poner solución a las limitaciones encontradas y mencionadas anteriormente, se propone un nuevo método para el desarrollo de datasets dedicados a clasificación y predicción estética de imágenes verificado con criterio humano bajo condiciones experimentales controladas. Este nuevo método engloba las ventajas de los conjuntos de imágenes estudiados previamente tras haber comprobado que su capacidad de generalización no es del todo satisfactoria.

La metodología desarrollada está basada en los cinco pasos que se muestran en la figura 5.2 y que se explican a continuación:

1. Obtención de una gran cantidad de imágenes evaluadas por diferentes usuarios. Se utiliza un proceso de fuerza bruta a través del que se descargan los datos de todas las imágenes.
2. Filtrado de las imágenes para que se cumplan unas condiciones mínimas. Se seleccionan aquellas imágenes que tienen más de 100 votos. De esta forma, se pretende que el valor medio de cada imagen se encuentre lo menos sesgado posible.
3. Clasificación de las imágenes en grupos en función de la puntuación media recibida en el portal de origen.

4. De cada uno de los grupos resultantes se selecciona un conjunto formado por un número determinado de imágenes con la menor desviación típica. Esto proporciona conjuntos con un número igual de elementos y alta coherencia en la puntuación de las imágenes, por lo que pueden ser representativas de un conjunto mayor.
5. El conjunto de imágenes resultante de los procesos anteriores es evaluado por un grupo de humanos en condiciones experimentales controladas.

5.3. Construcción de un nuevo dataset

A partir de la metodología diseñada y explicada en la sección anterior, se diseñó un nuevo conjunto de datos para la predicción y clasificación de imágenes, en el que el proceso de evaluación humana de las imágenes se llevó a cabo en condiciones experimentales controladas. Este nuevo método de creación permite construir un conjunto de imágenes con mayor coherencia estadística a partir de los resultados recogidos en la web de fotografía DPChallenge. Posteriormente, las imágenes seleccionadas se evalúan de forma similar al procedimiento utilizado por Forsythe et al. [49]. La creación de este dataset nos permite también analizar la correlación entre los resultados obtenidos con sujetos en condiciones controladas y los obtenidos a través del portal fotográfico.

5.3.1. Datos originales

La construcción del nuevo dataset comenzó con la recolección de un conjunto de imágenes del portal de fotografía DPChallenge. Las imágenes de este portal son calificadas por los usuarios dentro del rango [1, 10], donde 1 corresponde a la puntuación más baja y 10, a la más alta. La única información que existe sobre la puntuación en DPChallenge es que una imagen con una puntuación de 1 es una foto “mala”, y una imagen con una foto de 10 es “buena”, por lo que no se puede afirmar que la puntuación esté relacionada directamente con el valor estético, la calidad fotográfica o la originalidad. Sin embargo, este portal ha sido utilizado en estudios previos para obtener datos para experimentos de clasificación estética [87, 104, 154]. La idea original de este sitio web fue la de crear un lugar donde los amigos de amigos pudieran aprender a ser mejores fotógrafos a través de un *reto* cada semana. Metodológicamente, DPChallenge organiza concursos

Cuadro 5.2: Datos descriptivos del conjunto de datos original.

Datos descriptivos	
Imágenes	44.047
Media	5,2405
Desviación	0,7821
Varianza	0,6117
Curtosis	0,2480
Sesgo	-0,0182
Mínimo	1,9951
Máximo	8,3900

semanales estructurados por temas a partir de una palabra o frase (por ejemplo, “Alfred Hitchcock”, “Abstracto: blanco y negro” o “Retrato en color”). Para la creación del nuevo dataset no se tienen en cuenta estos temas.

Las imágenes se recogieron mediante un proceso de fuerza bruta en el que se descargaron todos los datos de las imágenes con identificadores comprendidos entre 10.000 y 172.000. Del total de 62.000 imágenes, solo estaba disponible toda la información estadística de las evaluaciones para 40.047 imágenes. Estas imágenes fueron valoradas por una media de 233 sujetos en DPChallenge y la valoración media fue de $5,23 \pm 0,78$. Todos los datos descriptivos se muestran en la tabla 5.2. El archivo con los datos de la evaluación y los enlaces a las imágenes utilizadas se pueden consultar de forma pública ¹. La figura 5.3 (a) muestra la disposición del número de votos dentro del rango de valoraciones válidas y la figura 5.3 (b) muestra la distribución de las evaluaciones medias de las imágenes dentro del rango de puntuaciones, resultando en ambos un modelo gaussiano.

¹<https://doi.org/10.6084/m9.figshare.6127295.v1>

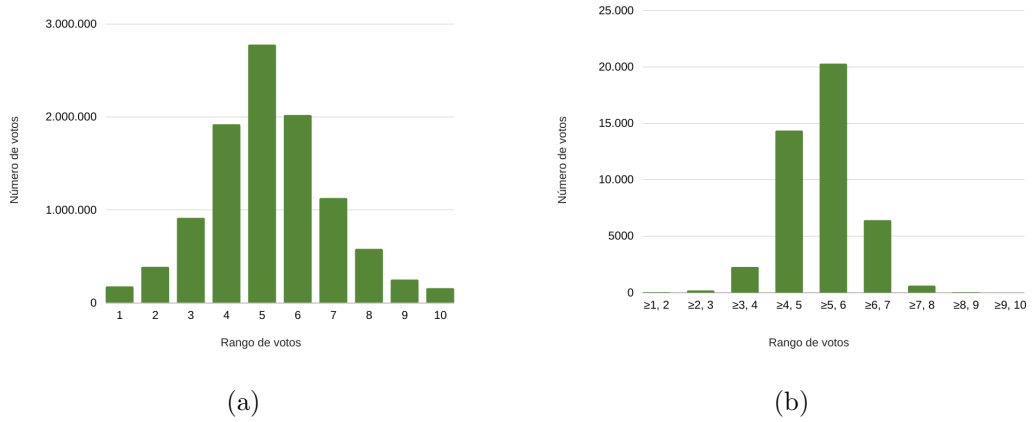


Figura 5.3: Caracterización de las 44.047 imágenes obtenidas inicialmente de DP-Challenge. (a) disposición del número de votos dentro del rango de valoraciones válidas y (b) distribución de las evaluaciones medias de las imágenes dentro del rango de valoraciones válidas.

5.3.2. Dataset propuesto

Como ya se mencionó en la sección anterior, solo se utilizaron las imágenes que contaban con todos los datos de evaluación. A continuación, se realizó otro cribado con la condición de que todas las imágenes tuvieran al menos 100 valoraciones. El objetivo de esta acción era que el valor medio de cada imagen fuera lo menos sesgado posible.

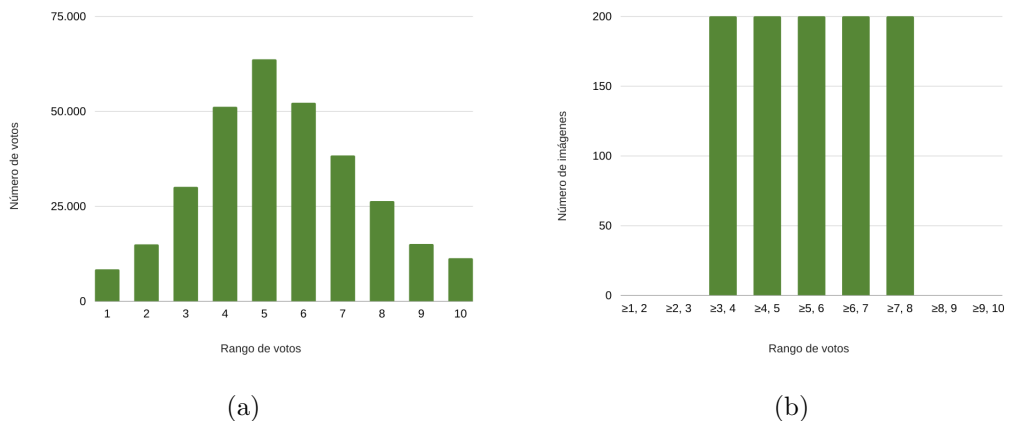


Figura 5.4: Caracterización de las 1.000 imágenes del conjunto propuesto. (a) Distribución del número de votos dentro del rango de puntuación y (b) distribución de las valoraciones medias en las imágenes dentro del rango de puntuación válido.

Una vez realizada esta selección, las imágenes se organizaron en grupos según las valoraciones medias obtenidas en DPChallenge. Las imágenes se clasificaron en 9 rangos de puntuación, uno por cada valor entero de evaluación válido. A continuación, se estableció un número mínimo de imágenes para todos los grupos. En este caso, el número de imágenes establecido como mínimo fue 200 (véase la figura 5.4 (b)). Con esta condición, no existían imágenes suficientes en los conjuntos de puntuación inferior a 3 ni en los de puntuación superior a 8. En consecuencia, los grupos utilizados para la construcción del dataset se encuentran en el rango de puntuación [3, 8]. De estos grupos, se seleccionaron las 200 imágenes con menor desviación estándar para el conjunto de evaluación final. En otras palabras, se seleccionaron las imágenes con las puntuaciones medias más consistentes. Los datos descriptivos de cada uno de los rangos se detalla en la tabla 5.3. La figura 5.4 muestra (a) la distribución del número de votos dentro del intervalo de puntuaciones válidas y (b) la distribución de las valoraciones medias dentro de las puntuaciones válidas para el conjunto de datos de 1.000 imágenes.

Este proceso proporciona un conjunto de imágenes con igual número de elementos para cada rango, con alta consistencia de puntuación, y que podría ser finalmente el más representativo.

Cuadro 5.3: Datos descriptivos de cada uno de los cinco conjuntos de 200 imágenes que componen el dataset propuesto.

Rango	[3,4)	[4,5)	[5,6)	[6,7)	[7,8)
Media	3,5943	4,4695	5,4975	6,4715	7,3112
Desviación	0,2613	0,2868	0,2894	0,2845	0,2335
Varianza	0,0683	0,0822	0,0837	0,0809	0,0545
Curtosis	-0,8224	-1,2370	-1,1765	-1,1595	-0,4500
Sesgo	-0,3998	0,1611	-0,0005	0,1099	0,6879
Mínimo	3,0070	4,0130	5,0060	6,0030	7,0000
Máximo	3,9970	4,9970	5,9970	6,9940	7,9530

5.3.3. Evaluación

El conjunto de imágenes descrito en la sección anterior fue evaluado por un grupo de humanos de forma presencial y en condiciones experimentales controladas.

Según Neyman [132], con un tamaño de muestra mínimo de 8 individuos y un

nivel de confianza del 95 %, se puede obtener la verdadera valoración poblacional de una imagen, con un margen de error del 3 %.

Para el proceso de evaluación, se crearon 5 subconjuntos con imágenes seleccionadas al azar del total de las 1.000 disponibles. Cada persona podía valorar las imágenes de uno o varios de estos subconjunto en una escala de Likert² de entre 1 y 5, siendo 1 la menor puntuación posible y 5, la mayor. Cada conjunto fue evaluado por al menos 10 personas (más de 10.000 valoraciones en total).

Para cada imagen, los usuarios calificaron independientemente su valor estético y su calidad. El texto de la encuesta decía: “En esta tarea queremos que evalúes la calidad y el valor estético de cada una de las imágenes que te proponemos. Para puntuar la calidad debes fijarte en aspectos como el encuadre, el enfoque o los colores. En general, las fotografías profesionales tienen mayor calidad que las realizadas por aficionados. La edición de las imágenes (uso de filtros, Photoshop, etc.) no tiene por qué afectar a su calidad. Puede ser que no te guste una imagen, pero si está bien hecha, su puntuación de calidad debería ser alta. Para la puntuación del valor estético, buscamos tu opinión personal sobre la imagen, si te gusta o no. El valor semántico no debe influir. Es decir, una imagen bonita de un bebé llorando puede tener una puntuación estética alta”.

Las evaluaciones presenciales fueron realizadas en dos sesiones por estudiantes voluntarios de la Universidade da Coruña (principalmente, estudiantes de la Facultad de Ciencias de la Comunicación). El número total de participantes fue de 99 (33 hombres y 66 mujeres), con una edad media de 18,7 en un rango total de edad [18, 30]. Cada participante evaluó al menos 200 imágenes y todos lo hicieron en las mismas condiciones de visualización: pantallas con las mismas especificaciones, mismas condiciones de iluminación y misma distancia entre los evaluadores y las pantallas.

Los datos que se muestran en la figura 5.5 corresponden a la media obtenida para cada imagen a partir de las diferentes valoraciones realizadas para los criterios de estética y de calidad.

²La escala de Likert [98] es una escala psicométrica que se utiliza para cuestionar el nivel de acuerdo o desacuerdo con una declaración. Se trata de uno de los tipos de escalas de medición utilizados principalmente en la investigación de ciencias sociales que emplean cuestionarios. Esta escala asume que la fuerza e intensidad de la experiencia es lineal, por lo que va desde un totalmente en desacuerdo a un totalmente de acuerdo. Las respuestas pueden ser ofrecidas en diferentes niveles de medición, permitiendo escalas de 5, 7 y 9 elementos. Siempre debe haber un elemento neutral para aquellos usuarios que no estén de acuerdo ni en desacuerdo.

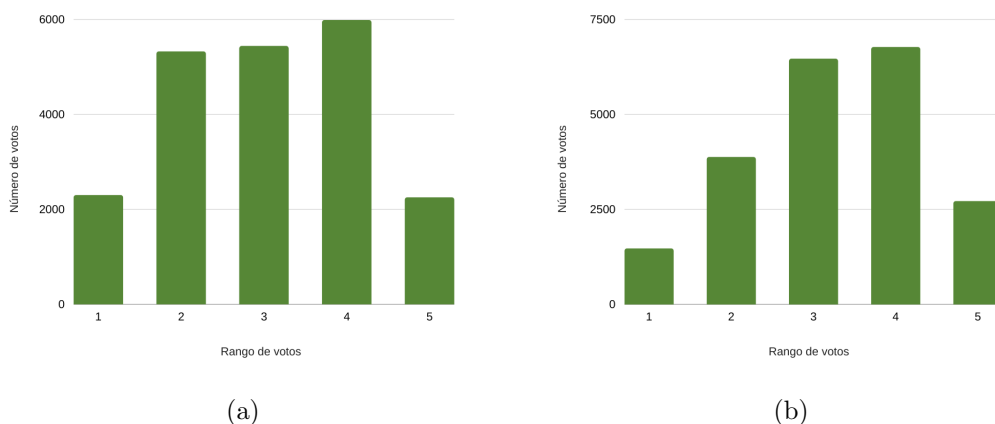


Figura 5.5: Distribución de las valoraciones medias de (a) estética y (b) calidad obtenidas en el grupo de control.

Se calculó la correlación entre las evaluaciones realizadas de forma presencial y las registradas en la plataforma DPChallenge (véase la figura 5.6). Se utilizaron las medidas de correlación Pearson y Spearman por ser dos de las más utilizadas en clasificación y predicción estética. La diferencia entre la medida de Pearson y la de Spearman reside principalmente en que la primera es independiente de la escala de medida de las variables, solo tiene en cuenta la expresión numérica de las mismas, mientras que la segunda requiere que las variables estén medidas en una escala ordinal y puede ser dependiente del orden de los datos.

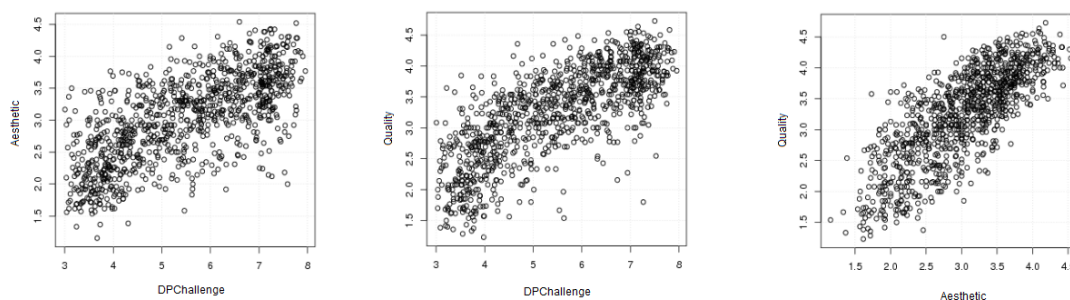


Figura 5.6: Gráficos de dispersión entre los rangos para las tres combinaciones posibles dados los criterios evaluados para todo el estudio.

La correlación de Pearson [137] entre la puntuación de DPChallenge y la puntuación media del valor estético fue de 0,692 y de 0,690 según el valor de Spearman. Por su parte, la correlación entre el valor de DPChallenge y la puntuación media de calidad fue de 0,748 según Pearson y de 0,756 según Separman.

Por último, la correlación entre las dos medidas obtenidas en el experimento presencial (estética/calidad) fue de 0,787 según Pearson y de 0,786 según Spearman, más alta que en las otras dos correlaciones. La figura 5.6 muestra los gráficos de dispersión entre los rangos para las tres combinaciones posibles teniendo en cuenta los tres criterios que se evalúan para todo el estudio.

En el apéndice B se puede encontrar una muestra de imágenes del dataset propuesto ordenadas por rangos según las tres puntuaciones recogidas (DPChallenges, calidad y estética).

5.4. Experimentación

En este estudio se propusieron algunos modelos de vanguardia basados en Machine Learning para su aplicación a la entrada propuesta en 5.3. El objetivo de estos experimentos es estudiar si los valores de correlación entre las evaluaciones (DPChallenge, estética y calidad) pueden ser replicados por un sistema informático para el conjunto de datos propuesto. Del mismo modo, se pretende validar la nueva metodología propuesta para la predicción del valor estético en imágenes digitales.

5.4.1. Materiales y métodos

Para caracterizar las imágenes que componen el conjunto de estudio ya descrito en la sección 5.3, se utilizó el extractor de características WND-CHARM [134], que es un clasificador de imágenes polivalente que puede aplicarse a una gran variedad de tareas de imagen. Según sus desarrolladores, el sistema extrae un amplio conjunto de características de la imagen, incluyendo descomposiciones polinómicas, características de alto contraste, estadísticas de píxeles y texturas, entre otras. Esas características se calculan sobre la imagen en bruto, sobre las transformaciones de la imagen y sobre las transformaciones de las transformaciones de la imagen. El vector de características final abarca 2.905 variables, cada una de las cuales informa sobre un aspecto diferente del contenido de la imagen. Todas las características se basan en imágenes en escala de grises, por lo que actualmente no se utiliza la información sobre el color.

Durante la fase de experimentación se probaron los diferentes modelos computacionales utilizando *10-fold cross-validation* para dividir los datos y 50 ejecuciones por modelo con el fin de evaluar el rendimiento en diferentes experimentos.

El rendimiento de los modelos se evaluó mediante el coeficiente de correlación de Spearman (ρ) y el coeficiente de correlación de Pearson (r de Pearson).

5.4.2. Modelos computacionales

Durante el proceso de experimentación se realizaron diferentes pruebas para seleccionar el mejor modelo usando el paquete R y MATLAB[®]. Algunos de los métodos computacionales utilizados buscaban el menor subconjunto de variables del conjunto original que proporcionara un mejor rendimiento [15], o al menos igual al obtenido al utilizar todas las variables posibles, considerando que se trataba de un enfoque de selección de características (Feature Selection) [16, 48, 143].

Concretamente, los métodos utilizados fueron los siguientes:

- Support Vector Machines-Recursive Feature Elimination (SVM-RFE) [66, 110]
- Generalized Linear Model con Stepwise Feature Selection (GLM) [72], que selecciona las características que minimizan la puntuación AIC y la Regresión Lineal Múltiple (LM) estándar más básica sin FS.
- k-Nearest Neighbors (k-NN) [42], concretamente la variante denominada k-NN ponderado.
- Generalized Boosted Models (GBM) [53].

Se mejoraron las capacidades del paquete RRegrs [163] para implementar el SVM-RFE y el GLM con el fin de evitar la búsqueda del mejor modelo según la metodología propuesta ya que, según García et al. [56], debe realizarse en base a una prueba de hipótesis nula. También se modificó este paquete para evitar el proceso de división inicial y se realizó un proceso de validación cruzada externa para evitar el sesgo de selección, tal como sugiere Ambroise y McLachlan [2].

El algoritmo del vecino más cercano (k-NN) es una técnica basada en la teoría de los *clusters*. En este caso, se utilizó una variante denominada k-NN ponderado [70]. Se basa en el hecho de que una nueva observación particularmente cercana a otra observación dentro del conjunto de aprendizaje debería tener una gran importancia en el proceso de toma de decisiones y, a la inversa, una observación que está a una distancia más lejana debería tener mucha menos importancia [101].

Para este algoritmo, solo se ajustó el hiperparámetro k , que representa el número de puntos de datos vecinos que se consideran más cercanos. El rango de valores fue de 1 a 5.

Los Generalized Boosted Models (GBM) aplicaron el enfoque descrito en [63] para establecer la base de los algoritmos de refuerzo. La estimación de GBM implica un proceso iterativo de regresión múltiple para captar relaciones complejas y no lineales sin sobreajustar los datos [53, 54]. Funciona con variables continuas y discretas y es invariable a sus transformaciones monotónicas [118]. Para este algoritmo, la profundidad interactiva se representó por el número de divisiones que tuvo que realizar en un árbol (partiendo de un solo nodo) y el número de árboles que se ajustaron. El rango de valores utilizados fue de 1 a 4 y de 100, 250 y 500 para el número de árboles.

El diseño de los experimentos realizados se basó en una metodología novedosa para el desarrollo de diseños experimentales en problemas de regresión con algoritmos de regresión de aprendizaje automático múltiple [48]. Para cada modelo descrito anteriormente, se buscó el conjunto óptimo de parámetros mediante la optimización de hiperparámetros.

5.4.3. Resultados

La figura 5.7 y la tabla 5.4 muestran los resultados obtenidos para cada uno de los cuatro modelos estudiados según el valor de correlación Pearson y Spearman, utilizando como referencia las valoraciones medias del portal DPChallenge. En primer lugar, examinando los valores de correlación de Spearman, el valor máximo del modelo basado en SVM fue de 0,574, con 1024 variables seleccionadas. El conjunto de entrada puede reducirse a 256 sin una pérdida significativa de rendimiento (0,570), ya que los valores de correlación se mantuvieron estadísticamente constantes entre ambas cifras. Por otra parte, si observamos los valores de la r de Pearson, se mantiene el mismo patrón, ya que con 1024 variables de entrada se obtuvo un 0,581, mientras que con 256 se obtuvo 0,574 (sin diferencia significativa de rendimiento). En cualquier caso, tanto los valores de Spearman como los de Pearson muestran una relación ascendente (positiva) moderada, con la excepción de k -NN, que presentan una relación descendente a medida que aumenta el número de variables de entrada.

Cuadro 5.4: Media de los resultados presentados en la figura 5.7, identificando los hiperparámetros y el tamaño de la entrada para cada modelo.

Tamaño	Modelo	Pearson	SD	Hiperparámetros
16	GLMNET	0,5320	0,0717	Alpha=0
16	GBM	0,5234	0,0738	Interaction.depth=4; n.trees=500
16	k-NN	0,4831	0,0744	k = 12; distance=2
16	SVM	0,5389	0,0713	Cost=16; Gamma=0,00984
32	GLMNET	0,5451	0,0709	Alpha=0
32	GBM	0,5266	0,0733	Interaction.depth=4; n.trees=500
32	k-NN	0,4851	0,0750	k=12; distance=2
32	SVM	0,5581	0,0732	Cost=0,397; Gamma=0,00984
64	GLMNET	0,5406	0,0669	Alpha=0
64	GBM	0,5474	0,0723	Interaction.depth=4; n.trees=500
64	k-NN	0,4898	0,0752	k=12; distance=2
64	SVM	0,5503	0,0691	Cost=2,52; Gamma=0,000244
128	GLMNET	0,5473	0,0745	Alpha=0
128	GBM	0,5425	0,0679	Interaction.depth=4; n.trees=500
128	k-NN	0,4926	0,0720	k=12; distance=2
128	SVM	0,5687	0,0676	Cost=0,397; Gamma=0,00155
256	GLMNET	0,5555	0,0719	Alpha=0,15
256	GBM	0,5479	0,0704	Interaction.depth=4; n.trees=500
256	k-NN	0,4776	0,0774	k=12; distance=2
256	SVM	0,5778	0,0671	Cost=2,52; Gamma=0,000244
512	GLMNET	0,5748	0,0701	Alpha=0,15
512	GBM	0,5482	0,0765	Interaction.depth=4; n.trees=500
512	k-NN	0,4845	0,0758	k=12; distance=2
512	SVM	0,5747	0,0683	Cost=2,52; Gamma=0,000244
1024	GLMNET	0,5644	0,0708	Alpha=0,15
1024	GBM	0,5473	0,0685	Interaction.depth=4; n.trees=500
1024	k-NN	0,4908	0,0777	k=12; distance=2
1024	SVM	0,5782	0,0670	Cost=0,397; Gamma=0,000244
2048	GLMNET	0,5602	0,0733	Alpha=0,15
2048	GBM	0,5465	0,0692	Interaction.depth=4; n.trees=500
2048	k-NN	0,4482	0,0815	k=12; distance=2
2048	SVM	0,5723	0,0685	Cost=2,52; Gamma=0,000244
fulldataset	GLMNET	0,5590	0,0719	Alpha=0,15
fulldataset	GBM	0,5476	0,0690	Interaction.depth=4; n.trees=500
fulldataset	k-NN	0,4299	0,0825	k=12; distance=2
fulldataset	SVM	0,5554	0,0721	Cost=2,52; Gamma=0,000244

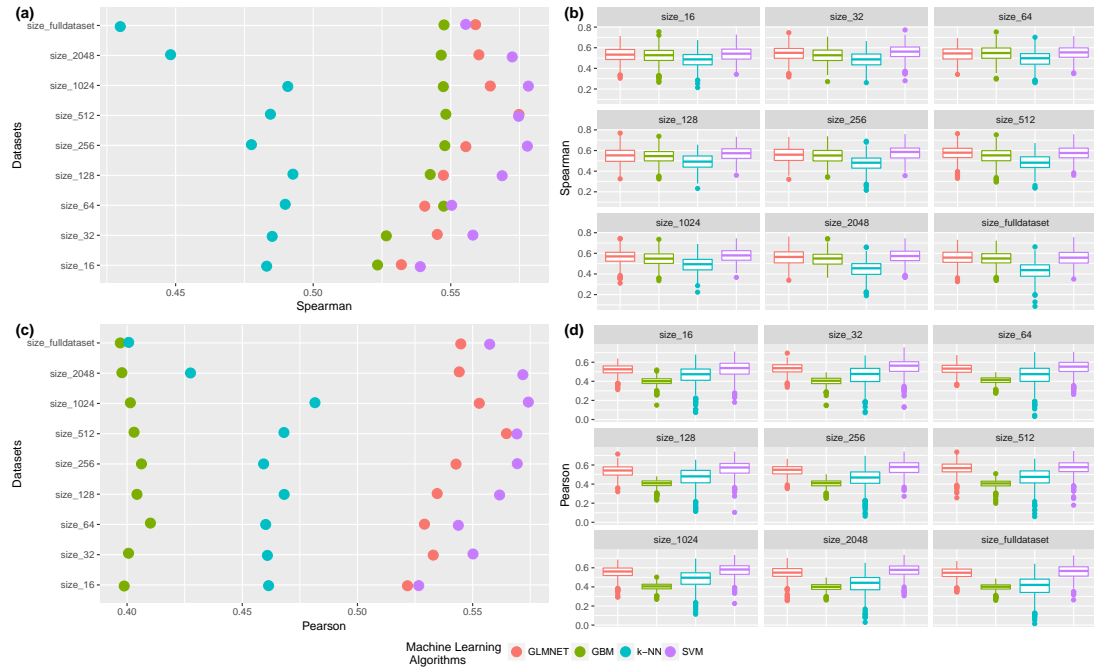


Figura 5.7: Resultados obtenidos para los cuatro modelos propuestos y optimizados mediante hiperparametrización. A la derecha, se muestran los valores medios de Spearman (arriba) y Pearson (abajo). A la izquierda, las distribuciones de las 50 ejecuciones independientes de cada modelo óptimo (Spearman arriba y Pearson abajo) con diferentes tamaños de entrada probados mediante FS.

Se comprobó la importancia de la diferencia entre GLM, SVM, GBM y k-NN con 256 variables de entrada (véase figura 5.8) mediante una prueba de Kruskal-Wallis [93], y los resultados revelaron que, con un nivel de confianza muy alto, SVM ($\text{cost} = 2^{-6}$ y $\text{gamma} = 2^{-9}$) era significativamente mejor que los demás con un valor $p < 2,2 \times 10^{-16}$. En consecuencia, se puede afirmar que el conjunto mínimo de entrada con los mejores resultados fue de 256 variables de entrada en combinación con un modelo de predicción SVM con parámetros especificados.

Una vez identificado el método con mejores resultados utilizando las valoraciones medias obtenidas por los usuarios de DPChallenge, se calcularon los mejores hiperparámetros SVM ($\text{cost} = 2^{-4}$ y $\text{gamma} = 2^{-12}$ en ambos casos) para entrenar con las puntuaciones de “estética” y “calidad” obtenidas en el experimento realizado con humanos.

Como se muestra en la figura 5.9, los valores de correlación para cualquiera de los tres casos son inferiores a 0,60 de media. En concreto, fue de 0,578 para DPChallenge, de 0,456 para el valor de estética y de 0,539 para calidad, utilizando como medida de rendimiento la rho de Spearman y 0,574, 0,451 y 0,562,

respectivamente, utilizando la r de Pearson.

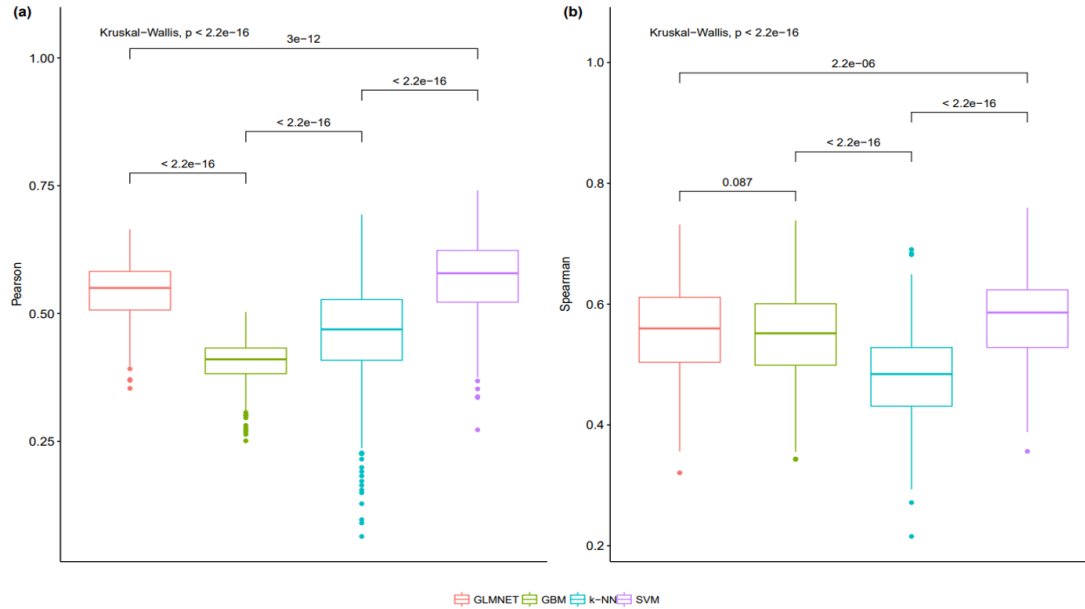


Figura 5.8: Distribución de las correlaciones obtenidas para cada modelo optimizado (Pearson a la derecha y Spearman a la izquierda). Para cada par, se muestra el valor p obtenido mediante una prueba de Kruskal-Wallis.

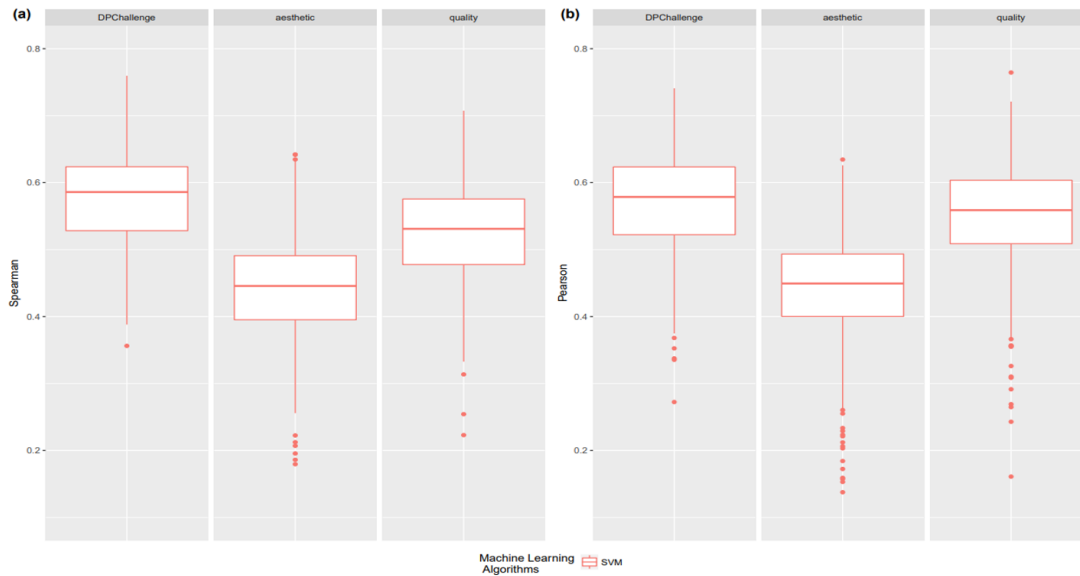


Figura 5.9: Distribución de las correlaciones (de Spearman a la derecha y de Pearson a la izquierda) obtenidas para cada uno de los tres criterios (DPChallenge, Estética y Calidad) utilizando 256 variables de entrada y un modelo SVM optimizado mediante hiperparametrización.

5.5. Discusión

Se obtuvo una correlación de 0,78 entre las valoraciones basadas en estética y las basadas en calidad. Esto indica que el grupo de evaluación distinguió mejor entre ambos criterios que en las mediciones realizadas por Datta et al. [35], donde la correlación de Pearson entre estética y originalidad fue de 0,891.

En cuanto a la correlación entre DPChallenge y calidad y estética individualmente, debemos empezar por destacar que la correlación más alta fue entre los valores de DPChallenge y los de calidad, lo que sugiere que los evaluadores de DPChallenge valoran más la calidad fotográfica que el valor estético de la imagen.

Desde una perspectiva más subjetiva, no existe una sola razón que explique la diferencia entre las correlaciones relativas a DPChallenge en lo que respecta a los valores de estética y calidad:

1. En el caso de DPChallenge, la valoración de los usuarios puede estar condicionada por la afinidad con el autor de la fotografía, ya que la valoración se realiza en formato concurso, mientras que en el caso del grupo presencial, las condiciones experimentales estaban controladas. Por ejemplo, todos los participantes utilizaban el mismo modelo de pantalla, a la misma distancia, con la misma luz, etc. En DPChallenge se pueden utilizar numerosos dispositivos para realizar la evaluación (*smartphones*, tabletas o pantallas de alta resolución) y las condiciones, como la distancia de visualización y la luz ambiental son heterogéneas.
2. En el caso del grupo presencial, los criterios de evaluación fueron establecidos: estética y calidad. En DPChallenge, como ya se ha mencionado, la evaluación se realiza con la premisa de un concurso de fotografía, por lo que los evaluadores pueden tener en cuenta diferentes aspectos como la calidad fotográfica, la originalidad o el gusto personal.
3. Por otro lado, en el caso de las valoraciones presenciales, el mínimo de evaluaciones por imagen era de 10 mientras que para las evaluaciones de DPChallenge el mínimo era de 100 por cada imagen. Hay que tener en cuenta que para el conjunto de DPChallenge se seleccionaron las imágenes con menor desviación estándar, lo que supone que la nota media de DPChallenge tenía una desviación estándar de 0,27, mientras que las valoraciones de estética y calidad realizadas de forma presencial contaban con una desvia-

ción estándar de 1,18 y 1,10, respectivamente. Este puede ser el motivo de que SVM presente un valor de correlación mayor para DPChallenge, que para estética y calidad.

Al observar las características visuales de algunas de las imágenes del conjunto (figura 5.10), se encontraron algunos casos que merecen un análisis específico:

1. La figura 5.10(a) fue penalizada por los usuarios del portal de fotografía por tener algunas zonas sobreexpuestas. Muestra una palmera en primer plano que está ligeramente mal expuesta. Sin embargo, obtuvo una puntuación alta en estética porque tenía cierto valor estético para los evaluadores. El valor de calidad fue más cercano al de DPChallenge en este caso.
2. La figura 5.10(b) obtuvo también una calificación baja en DPChallenge, mientras que en calidad y estética se encuentran claramente por encima de la media. Esta diferencia podría deberse a las condiciones experimentales en las que tuvo lugar la evaluación presencial (buena calidad de imagen en una pantalla con buena resolución, donde se podría ver el cielo correctamente expuesto). En estas condiciones, los evaluadores podrían haber prestado más atención a la gota y al cielo en detrimento de la zona subexpuesta.
3. La figura 5.10(c) en DPChallenge tuvo una alta calificación, lo que puede deberse a que se tuvieron en cuenta criterios como la originalidad y la calidad de la edición.
4. Por último, en la figura 5.10(d) la calidad fue de nuevo más cercana al valor de DPChallenge. Sin embargo, obtuvo una puntuación más baja en estética

Todo lo anterior demuestra que en DPChallenge, en casos concretos, se pueden estar valorando criterios muy diferentes: originalidad, calidad, estética, edición fotográfica, etc.

En cuanto al uso de técnicas de aprendizaje automático para predecir cada uno de los tres criterios estudiados, la mayor correlación obtenida fue de 0,578 utilizando SVM. Este valor es similar a los obtenidos por Marin y Leder [115] utilizando como criterios *arousal* (rho de Spearman = 0,44) y *pleasantness* (rho de Spearman = 0,64) o por Nadal et al. [128] con *beauty* (rho de Spearman = 0,648) en condiciones experimentales similares a las adoptadas en este experimento. Estos valores se obtuvieron utilizando numerosos métodos del estado del arte

para la predicción y determinación de la mejor configuración para cada uno de ellos mediante la hiperparametrización.

En cuanto a los valores de correlación individuales de calidad y estética utilizando SVM, se deduce que para el sistema fue más sencillo aprender los valores de calidad que los de estética, lo que cobra sentido si se tiene en cuenta que el primero es un componente menos subjetivo y más relacionado con las características intrínsecas de la imagen.

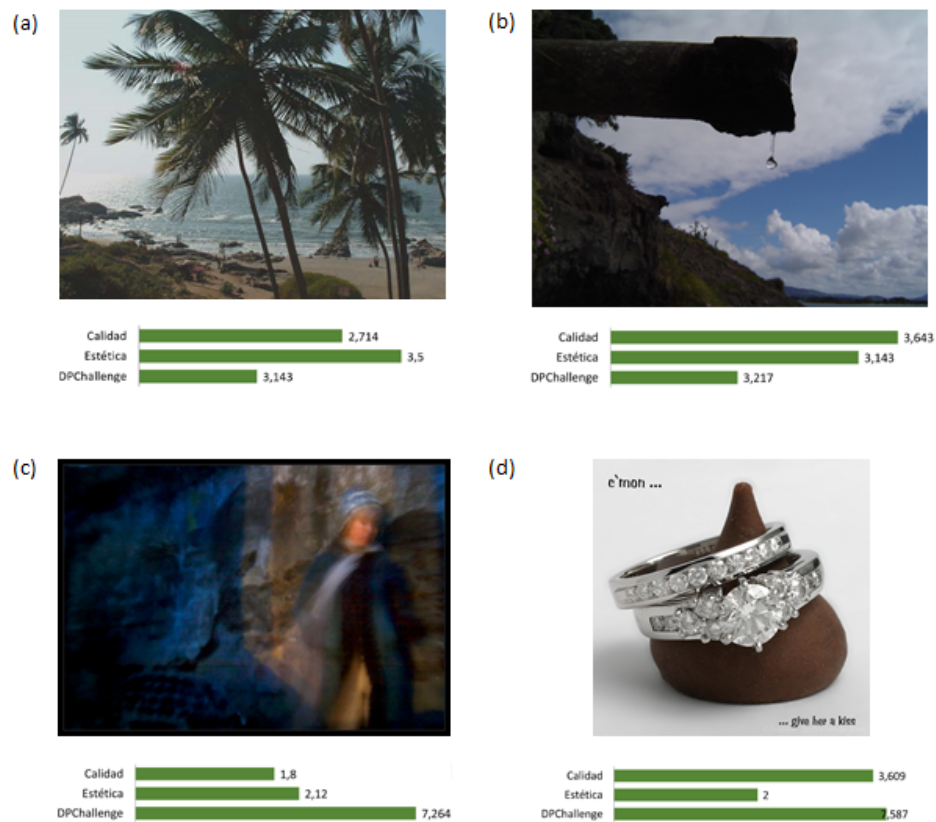


Figura 5.10: Ejemplos de imágenes con diferentes puntuaciones basadas en los tres criterios de evaluación. Para cada imagen, se da un valor numérico según cada criterio, mientras que las barras muestran el peso normalizado de dicho valor dentro de cada rango de evaluación (DPChallenge en el rango [1, 10] y estética y calidad en el rango [1, 5]).

5.6. Conclusiones

Teniendo en cuenta una serie de problemas encontrados en relación con los datasets del estado del arte, se desarrolló un nuevo conjunto de datos a partir de una nueva metodología diseñada. Este conjunto de datos consta de 1.000 imágenes del portal fotográfico DPChallenge, que se evaluaron de tres formas diferentes: (i) evaluación a través del portal fotográfico DPChallenge con al menos 100 valoraciones por imagen; (ii) evaluación del valor estético realizada en condiciones experimentales controladas y con un mínimo de 10 votos por imagen; y (iii) evaluación de la calidad de las imágenes realizada en las mismas condiciones que (ii).

Los resultados de correlación sugieren que la evaluación de DPChallenge se acerca más a un criterio de calidad que a un criterio estético. Los usuarios de DPChallenge y los evaluadores de la prueba presencial califican las imágenes de forma diferente y es evidente que en DPChallenge cada usuario puede seguir criterios diferentes para la evaluación de las imágenes, como la originalidad, la calidad o el gusto personal.

Se utilizaron diferentes técnicas computacionales de última generación, se identificaron sus configuraciones óptimas y se aplicaron los tres criterios (DPChallenge, estética y calidad), con lo que se obtuvieron correlaciones de 0,578, 0,456 y 0,539, respectivamente. Estos resultados son similares a los obtenidos en los experimentos del estado del arte y muestran que las técnicas de aprendizaje automático tienen una capacidad mayor de aprender la percepción humana de la calidad técnica que otros criterios más subjetivos como la estética, a pesar de que la diferencia entre ambos es muy estrecha.

Capítulo 6

Comparación de modelos computacionales para la predicción estética

En este capítulo se utiliza un nuevo algoritmo genético híbrido para la predicción del valor estético de imágenes digitales y se compara con otros modelos de vanguardia, incluidos los dos que obtuvieron mejores resultados en el capítulo anterior: SVM y GLMNET.

6.1. Introducción

Como se comentó en el capítulo 3 del estado del arte, es habitual que para los problemas de predicción estética, sean los investigadores los que crean métricas ad-hoc basándose en la suposición de que dichas métricas están vinculadas al valor estético. Este trabajo aporta un nuevo enfoque que no depende de la definición de métricas por parte del investigador, sino que hace uso del aprendizaje por transferencia para esta tarea. Este enfoque automático mejora significativamente los resultados del estado del arte.

El aprendizaje por transferencia es un método que trata de reutilizar la representación de características aprendida de un problema para resolver otro, extrayendo características de una sola capa, cerca del final de la red, y alimentando a otro clasificador. Este enfoque se ha utilizado con éxito anteriormente para la clasificación [11, 41, 97]; sin embargo, hasta donde se tiene conocimiento, no se ha aplicado para tareas de regresión. En este trabajo, este enfoque se utiliza

junto con la representación interna de CNNs diseñadas para detectar objetos en imágenes como fuente de métricas para clasificar y predecir tareas. En concreto, se utiliza la capa interna de GoogleNet y ResNet.

Se ha utilizado un novedoso modelo híbrido CNN Correlation by Genetic Search (CGS) para la evaluación de la estética. Este sistema se basa en un algoritmo genético cuyo objetivo no es solo determinar las métricas más importantes, sino también crear un modelo basado en una regresión múltiple que maximice la correlación (Spearman, Pearson o R-cuadrado) modificando el espacio de búsqueda según los valores seleccionados durante el filtrado. Este enfoque parece ser muy eficaz con poblaciones pequeñas, es decir, con poblaciones inferiores al doble de la longitud de representación de los individuos. En el caso de poblaciones de mayor tamaño, el coste computacional se vuelve exponencial y no hay garantía de que el resultado corresponda a un mínimo local.

CGS fue entrenado con el dataset creado por Datta et al. [35] con origen en la web Photo.net, ya analizado en el capítulo 4 del estado del arte. Se eligió este conjunto de datos por ser uno de los más referenciados tanto en clasificación como en predicción estética.

Para evaluar mejor la eficacia de CGS, los resultados se comparan con los resultados de los métodos más utilizados en el estado del arte, utilizando en todos los casos las características de GoogleNet y ResNet.

6.2. Trabajos relacionados

Aunque existen numerosos trabajos de investigación que utilizaron el conjunto de datos obtenidos de Photo.net, el enfoque de este análisis son aquellos trabajos que tienen mejoras significativas en términos de estándares de calidad desde que el conjunto de datos fue publicado por primera vez.

En 2006, Datta et al. [35] emplearon el color, la textura, la forma, la composición y características ad-hoc de alto nivel para extraer ciertas características visuales, basándose en la suposición de que pueden discriminar entre imágenes estéticamente agradables y desagradables. Construyeron clasificadores automatizados utilizando SVM, árboles de clasificación y regresión lineal sobre términos polinómicos de las características y se aplicaron para inferir las valoraciones estéticas obtenidas por imágenes recogidas del portal de fotografía Photo.net. Obtuvieron una precisión de clasificación del 68,77% y una r de Pearson = 0,689.

En 2009, Wong y Low [170] presentaron un método mejorado por saliencia para la clasificación de fotografías e instantáneas profesionales que extrae las regiones salientes de una imagen utilizando un modelo de saliencia visual. Extrajeron un conjunto de características salientes que caracterizan al sujeto y describen la relación sujeto-fondo. Su enfoque produjo una precisión de clasificación de validación cruzada con 5 repeticiones (5-CV) del 78,8 % también utilizando SVM.

Marchesotti et al. [114] no propusieron ninguna característica específica, sino que adoptaron múltiples descriptores de imagen SIFT para predecir la calidad de las fotografías, así como un descriptor para capturar el color. Su método basado en SVMs lineales con pérdida utilizando la formulación primal y un algoritmo de descenso de gradiente estocástico (SGD) obtuvo una precisión de clasificación del 75,85 %.

En 2016, Wang et al. [167] propusieron un conjunto de características que incluía 16 características nuevas y 70 características probadas anteriormente. Con este conjunto de características, utilizaron un esquema de aprendizaje automático que consistía en un clasificador basado en SVM para estimar si una imagen tenía un valor estético alto o bajo. Según los autores, los experimentos realizados con este conjunto de datos arrojaron una precisión de clasificación del 82,24 % y un buen rendimiento comparable a la calificación estética humana, con un coeficiente de correlación Pearson de 0,790.

En 2017, Xia et al. [172] introdujeron un novedoso modelo que evaluaba la calidad de los medios de comunicación imitando el proceso humano que ve activamente una imagen. En primer lugar, construyeron el espacio perceptual que captura sin problema los *graphlets* tanto a bajo como a alto nivel. A continuación, un algoritmo de factorización de matrices no negativas (NMF) con preservación de la localidad extrajo las trayectorias de desplazamiento de la mirada (GSP) de cada medio. Por último, los GSP se utilizaron para aprender un modelo probabilístico de clasificación de la calidad de los medios, con una precisión del 86,14 %.

La tabla 6.1 muestra los valores de *accuracy* (cercanía de un valor medio a un valor estándar conocido [157]), AUROC (comparación de la sensibilidad frente a la especificidad en un rango de valores para la capacidad de predecir un resultado dicotómico [167]) y el valor de correlación Pearson (medida de la fuerza de la relación lineal entre dos variables de este tipo [137]). La última fila de la tabla presenta el valor máximo conocido actualmente.

Tanto el *accuracy* como la r de Pearson son las medidas más utilizadas en

Cuadro 6.1: Métodos y resultados del estado del arte en clasificación y predicción de estética usando el dataset de Photo.net [35] y empleando *accuracy*, AUROC y el valor de correlación Pearson como estándares de calidad.

Autores	Features	Método	Accuracy	AUROC	Pearson
Datta et al. [35]	15	SVM RBF	0,6877	0,8089	0,6890
Wong y Low [170]	15	SVM RBF	0,7467	0,8590	-
Marchesotti et al. [114]	224	SVM SGD	0,7585	-	-
Wang et al. [167]	86	SVM RBF	0,8240	0,8956	0,7900
Tan et al. [153]	56	Autoencoder	0,8450	0,8670	0,8430
Xia et al. [172]	56	GSP-GMM	0,8614	-	-
Valor máximo			0,8614	0,8956	0,8430

este campo para medir el rendimiento en problemas de clasificación y predicción, respectivamente. Dado que *accuracy* es una medida no robusta porque no tiene en cuenta más que los aciertos del modelo, se ha optado por proporcionar también el valor AUROC. Esta es una medida robusta y se obtiene de forma combinada teniendo en cuenta los errores de tipo I y de tipo II. Esta medida permite realizar una comparación objetiva entre diferentes modelos.

6.3. Algoritmo híbrido de aprendizaje automático

En este trabajo, la evaluación estética de las fotografías (clasificación y predicción) se realiza a través de un algoritmo genético híbrido que ha sido utilizado con éxito previamente para otras tareas de predicción [136].

6.3.1. Características de aprendizaje por transferencia

En primer lugar, se generaron vectores de características a partir de imágenes empleando *deep learning*. Las redes neuronales han demostrado ser un método basado en Inteligencia Artificial con numerosas utilidades, desde la segmentación de imágenes para radares [165], hasta la identificación de casos de esclerosis múltiple en medicina [166], la estimación de la temperatura diaria del suelo a distintas profundidades [86] o la previsión del caudal de los ríos [155]. Las redes neuronales

convolucionales (CNN) [92] son la técnica más avanzada para la clasificación de imágenes, pero su uso puede ser inviable en muchas ocasiones, ya que requieren conjuntos de entrenamiento muy grandes (desde miles hasta varios millones de imágenes) [147]. El enfoque en el estado del arte para superar esta dificultad consiste en aplicar el aprendizaje por transferencia (*transfer learning*) [147], una técnica que reutiliza parcialmente un modelo entrenado en una tarea de origen para una nueva tarea de destino [31, 60, 122, 135].

Como explican Sharif Razavian et al. [147], el aprendizaje por transferencia puede utilizarse de diferentes formas. El enfoque seguido en este trabajo consiste en utilizar la salida de la red entrenada en el origen como características “listas para usar” que se emplean para entrenar un nuevo clasificador para la tarea objetivo [135]. Con el fin de aplicar el aprendizaje por transferencia en nuestro contexto, se consideraron dos CNNs disponibles públicamente: GoogleNet [151] y ResNet 50 [69]. Estas dos redes se entrenaron inicialmente para la tarea de clasificación de imágenes del reto ImageNet Large Scale Visual Recognition Challenge (ILSVRC challenge) [142], un conjunto de datos de 1,2 millones de imágenes etiquetadas a mano con 1.000 categorías.

Para utilizar GoogleNet y ResNet como extractores de características de una imagen, se eliminó la última capa de estas redes entrenadas, se cambió el tamaño de la imagen (231 x 231 para GoogleNet y 224 x 224 para ResNet), se alimentaron las redes con las imágenes y se utilizó la salida producida por la penúltima capa de las redes como vector de características. De esta forma, se obtuvo un vector de 1.000 características en el caso de GoogleNet y uno de 2.048 en el caso de ResNet.

El algoritmo genético trabaja sobre ambos datos de entrada (características de GoogleNet y ResNet) e intenta maximizar el valor de correlación de Pearson.

Este procedimiento se implementó en Python utilizando OpenCV [82] para el procesamiento de las imágenes y el framework Keras [30] con un *back-end* de Tensorflow [1] para extraer las características de las CNN.

6.3.2. Método híbrido supervisado de aprendizaje automático (CGS)

La evaluación estética de las imágenes se trata habitualmente como un problema de aprendizaje automático. El proceso de aprendizaje automático se basa en la correspondencia de las características de la imagen con categorías o valo-

res estéticos. Debido a la subjetividad intrínseca del valor estético, los métodos de aprendizaje supervisado son los más populares, y muchos estudios anteriores utilizaron SVM con buenos resultados, demostrando así su eficacia.

El grupo de investigación que respalda esta tesis doctoral, desarrolló un algoritmo genético híbrido [117] basado en modelos de regresión (CGS - Correlation by Genetic Search), que se utilizó previamente en [136] para predecir el crecimiento urbano vertical en Tokio utilizando datos macroeconómicos. La selección de características [94], la transformación de características [99] y la selección de parámetros [76] se realizan simultáneamente para crear un modelo de regresión lineal ajustado utilizando el valor de correlación de Pearson [137] como medida de rendimiento en el proceso evolutivo.

El algoritmo genético utiliza operadores matemáticos para refactorizar las variables de entrada con el fin de encontrar una solución adecuada. Las 12 transformaciones disponibles en el proceso evolutivo utilizadas se representan en los diagramas de la figura 6.1.

El algoritmo genético intenta calcular la mejor combinación posible de selecciones y transformaciones para todas las características de entrada. El flujo de trabajo del algoritmo genético híbrido descrito anteriormente se encuentra en el diagrama de la figura 6.2. A través del proceso evolutivo, se determinan las mejores combinaciones de transformaciones en las variables de entrada, que maximizan la función objetivo previamente determinada.

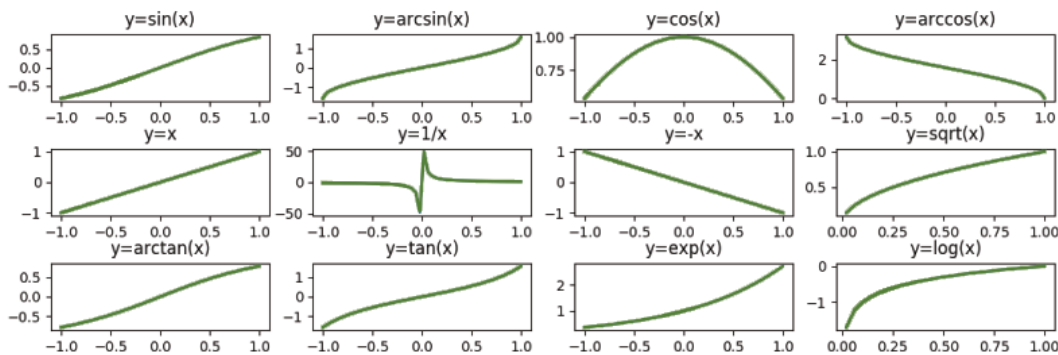


Figura 6.1: Las 12 transformaciones matemáticas disponibles que se utilizaron para determinar el proceso evolutivo.

Durante el proceso de selección de características, se conoce la posibilidad de determinar la utilidad de una variable al asignar la selección nula (figura 6.2). Para los mejores individuos de la población, se seleccionan los parámetros

con mejor adaptación (selector de parámetros) para crear el modelo de regresión final. Los criterios de terminación del proceso evolutivo están relacionados con el rendimiento de los individuos actuales de una población en relación con el individuo medio de la población. Cuando la diferencia media es inferior al valor umbral preestablecido para una población homogénea, el proceso de búsqueda iterativa llega a su fin. Para los casos en los que no se alcanza el valor umbral, se utiliza un número máximo de iteraciones para finalizar el proceso.

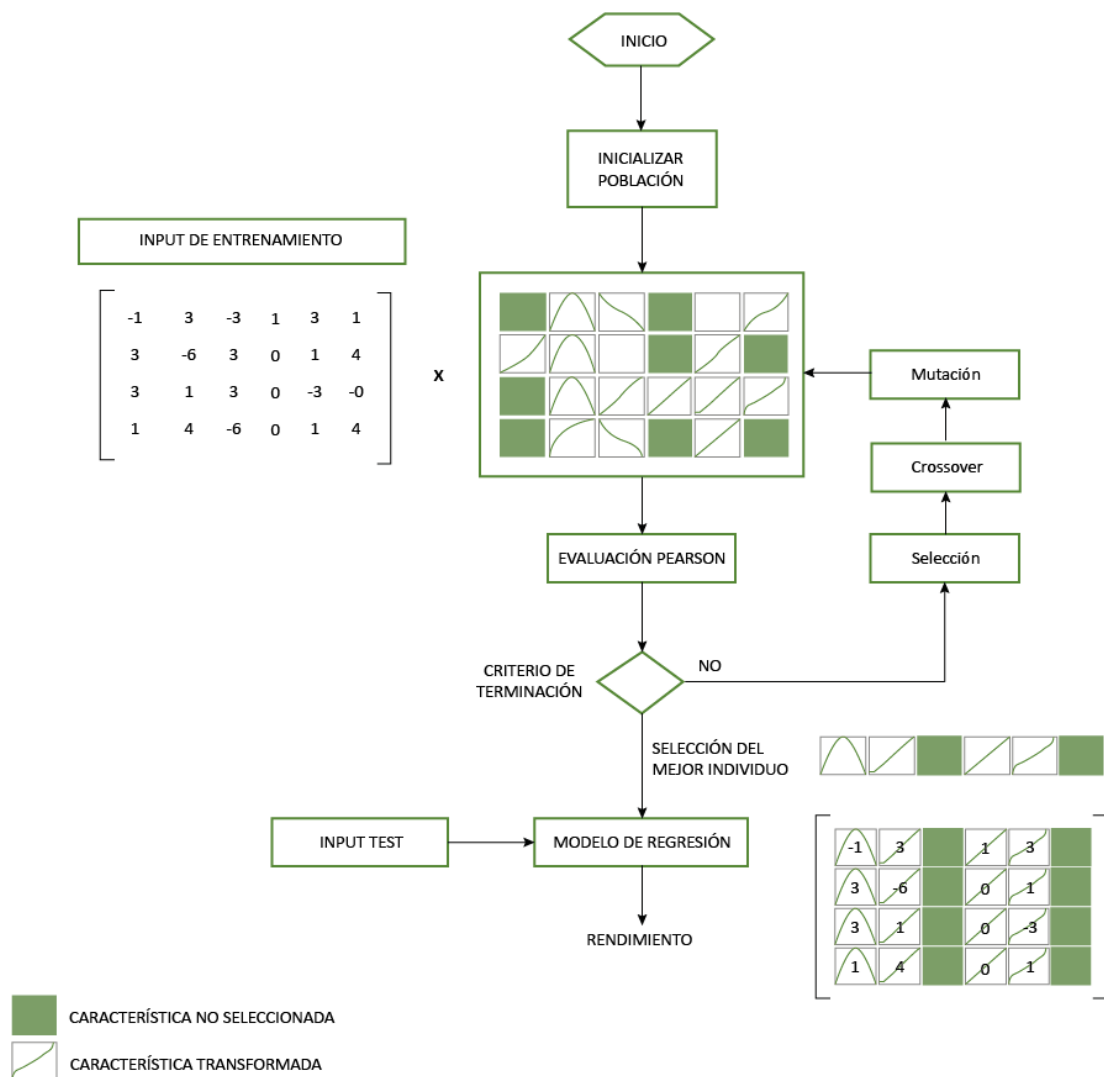


Figura 6.2: Diagrama de flujo de trabajo del algoritmo genético híbrido.

En este caso, el algoritmo genético trabaja sobre ambos datos de entrada (características de GoogleNet y ResNet) e intenta maximizar el valor de correlación de Pearson. Dado que la clasificación binaria es uno de los problemas tratados

en este trabajo, se modificó el modelo presentado en [136] siguiendo un modelo basado en la clasificación mediante regresión [51].

La razón de adoptar este enfoque radica en el propio método de regresión. El modelo híbrido está orientado a maximizar la función objetivo (en este caso el valor de regresión de Pearson) en lugar de minimizar el error de regresión asociado, como otros métodos, por lo que permite intrínsecamente “ignorar” aquellas entradas que el sistema identifica como atípicas. Dado que el error se ve afectado en mayor medida por el número de valores atípicos en comparación con otras medidas como el valor de Pearson, es razonable suponer que el uso de un método de clasificación por regresión en este caso puede ofrecer resultados satisfactorios.

6.4. Resultados

Esta sección valida el rendimiento del algoritmo genético híbrido en dos experimentos, uno de predicción y otro de clasificación, utilizando dos conjuntos de características de entrada diferentes (GoogleNet y ResNet).

Se realizó una validación cruzada de diez repeticiones (10-CV) para verificar la precisión del modelo en 50 experimentos independientes. Así, la validación cruzada puede evitar eficazmente cualquier problema de sobreajuste y dar una idea de la capacidad de generalización del modelo. La validación cruzada consiste en dividir la muestra de datos en subconjuntos independientes, realizar un análisis en uno o varios subconjuntos y validar el análisis en los subconjuntos restantes para reducir la variabilidad y detectar problemas como el sobreajuste.

Se obtuvieron los resultados de los 500 experimentos diferentes (50 ejecuciones distintas por cada combinación de validación cruzada).

En la figura 6.3 se muestra un resumen general de los resultados. Todos los resultados superan significativamente los mejores resultados del estado del arte. En el experimento de predicción, se obtuvieron correlaciones de Pearson de 0,9378 con GoogleNet y de 0,9177 con ResNet. En el experimento de clasificación, CGS con GoogleNet obtuvo un valor AUROC de 0,9293 y un *accuracy* de 0,9162.

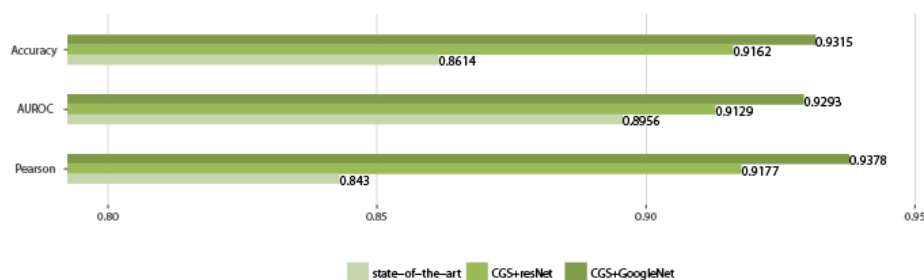


Figura 6.3: Resultados experimentales de ambos modelos híbridos de aprendizaje automático CNN-CGS comparados con el estado del arte.

Los mejores resultados de en la figura 6.3 se alcanzaron con los parámetros que se muestran en la tabla 6.2.

Cuadro 6.2: Parámetros con los que se encontraron los mejores resultados de CGS con GoogleNet y ResNet.

	GoogleNet	ResNet
Individuos	100	200
Epochs	900	700
Max threshold	0:001	0:001
Crossover rate	0:99	0:98
Mutation rate	0:01	0:02

Una de las funciones del CGS es filtrar las características derivadas de las CNN. Llama la atención que, a pesar de que GoogleNet proporciona menos características que ResNet (1.000 frente a 2.048), CGS incorpora más características de GoogleNet (274,5) que de ResNet (260,4). Esto puede deberse a que las características proporcionadas por ResNet están relacionadas con los valores residuales, con una alta correlación entre ellos. Así, el porcentaje de características filtradas es mayor; es decir, hay menos características que ofrezcan información útil al sistema CGS en ResNet.

Una de las razones por las que este enfoque puede superar los resultados del estado del arte es que, al utilizar un conjunto de métricas no diseñadas por humanos, es posible captar características de la imagen en las que nadie pensó antes, pero que pueden ser importantes para el trabajo de predicción y clasificación

estética. Además, el uso de CGS permite una exploración más completa de las métricas que permite detectar métricas eficaces para las tareas de predicción y clasificación.

6.4.1. Comparativa con los métodos del estado del arte

Para evaluar de forma más exhaustiva la eficacia de CGS, en esta sección se comparan los resultados de CGS con los resultados de los métodos más habituales del estado del arte, utilizando en todos los casos las características de GoogleNet y ResNet.

Como se ha comentado en los apartados anteriores, en general, los métodos de clasificación y predicción más utilizados en el estado del arte son los basados en SVM-RBF. Los dos conjuntos de características empleados en este trabajo (GoogleNet y ResNet) han demostrado obtener mejores resultados con esta técnica utilizando la hiperparametrización, tal como afirman Wang et al. [167]. Además, también se utiliza frecuentemente Random Forest (RF) [17] y el modelo lineal generalizado con selección de características por pasos (GLMNET) [72]. Los algoritmos de aprendizaje automático tienen diferentes parámetros que deben ajustarse para obtener el mejor rendimiento. Por ello, una de las fases típicas al trabajar con este tipo de algoritmos es la búsqueda en el espacio de la mejor combinación de hiperparámetros de cada modelo en particular. Esta exploración del espacio de soluciones permite identificar cómo afecta la variación de cada uno de dichos hiperparámetros al rendimiento del algoritmo, así como cuantificar la importancia de cada hiperparámetro en la solución final o en la convergencia del algoritmo. En este trabajo, el espacio de búsqueda de cada uno de los algoritmos utilizados se definió de acuerdo a lo indicado en el estado del arte para una búsqueda en cuadrícula simple [59].

RF Y GLMNET son capaces de discriminar las variables más relevantes durante la fase de entrenamiento de los modelos, por lo que pueden descartar las menos importantes. Estos procesos se realizan de forma diferente para cada uno de los algoritmos. Además, GLMNET es uno de los algoritmos de regulación por el que se permite variar las restricciones del modelo y su simplificación. Otras opciones posibles serían las de SVM, K-NN, o Naïve Bayes, pero ninguno ofreció resultados comparables.

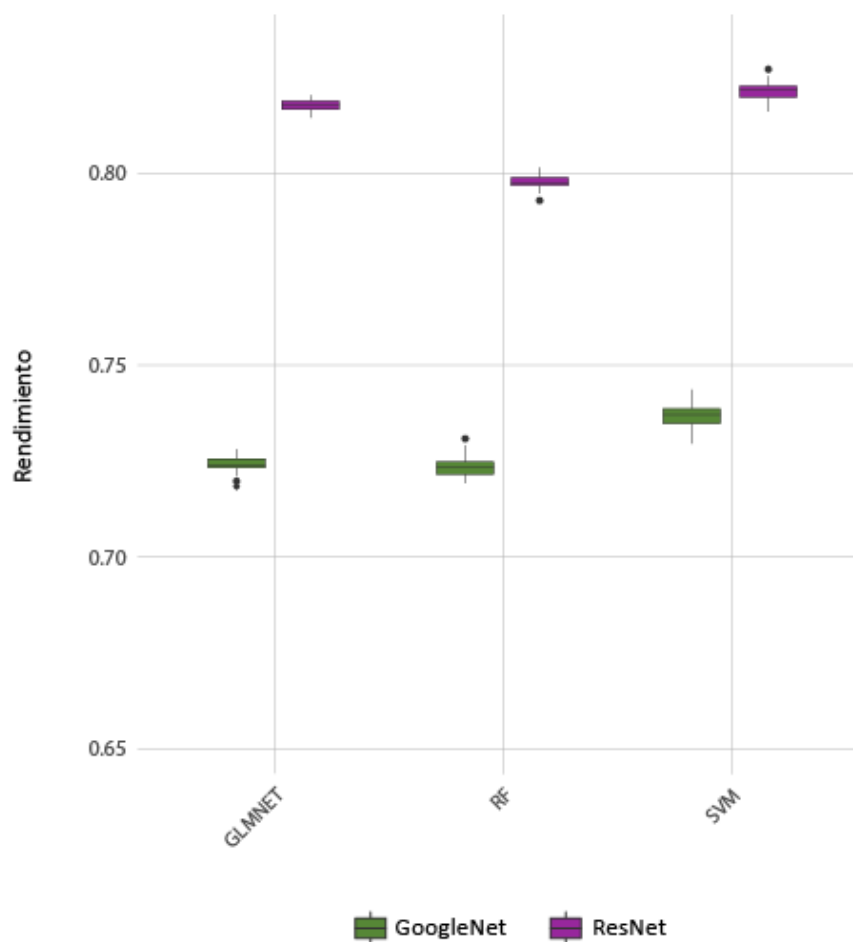


Figura 6.4: Resultados de 50 ejecuciones independientes en AUROC para GLMNET, Random Forest y Support Vector Machines con GoogleNet (verde) y ResNet (violeta)

Se realizaron 50 experimentos independientes para cada uno de los métodos del estado del arte y se utilizó AUROC para medir el rendimiento. Los resultados de GoogleNet se muestran en la figura 6.4 en color verde y los de ResNet en color violeta. Los resultados de AUROC siguen estando lejos de los del estado del arte que se muestran en la figura 6.3. Parecen indicar la idoneidad de CGS para estos vectores de características CNN.

Los mejores resultados en la figura 6.4 se alcanzaron con los parámetros mostrados en la tabla 6.3.

Cuadro 6.3: Parámetros con los que se encontraron los mejores resultados para los métodos del estado del arte utilizando GoogleNet y ResNet.

Método	Parámetro	GoogleNet	ResNet
GLMNET	α	0	0
RF	mtry	9	11
SVM	cost	16	2
	gamma	212	212

- Para GLMNET, se evaluó ElasticNet, mezclando el parámetro α con valores que oscilan entre 0 y 1, donde $\alpha = 1$ es la penalización del *lasso*, y $\alpha = 0$ la penalización de la *ridge*, utilizando el enfoque de rejilla simple.
- Para SVM, se evaluaron los hiperparámetros C y $gamma$ para un kernel RBF con valores que varían las potencias de 2, de -12 a 12.
- Para RF, se evaluó el número de variables disponibles para la división en cada nodo del árbol, hiperparámetro con valores que van de 5 a la raíz cuadrada del número de características.

Como se muestra en la figura, una vez identificados los mejores parámetros, SVM ofrece un mejor rendimiento para ResNet y GoogleNet (valores AUROC de 0,82 y 0,74) en comparación con GLMNET (valores AUROC de 0,81 y 0,72) y RF (valores AUROC de 0,79 y 0,72). Es importante destacar que mientras los mejores resultados de CGS utilizan la CNN GoogleNet, los mejores resultados de SVM, RF y GLMNET utilizan ResNet. En estos experimentos con GoogleNet solo se utiliza la última capa de representación, con 1.000 características frente a las 2.048 de ResNet. Los resultados con GLMNET, RF y SVM pueden mejorar con la información de otras capas.

Se puede considerar que el método CGS filtra eficazmente los conjuntos de entrada aunque los datos estén muy correlacionados. En otros métodos, como el RF, puede darse el caso de que dos estructuras de árbol utilicen variables diferentes, pero que tengan una correlación prácticamente total entre ellas, no ofreciendo información nueva que sea relevante para la clasificación o la regresión. Con CGS, cada variables solo se puede utilizar una vez de forma independiente. Además,

en CGS se aplica un operador a cada variable de entrada, transformando así el espacio de entrada. Así, dos variables que están muy correlacionadas inicialmente pueden no estarlo una vez realizada esta transformación.

Además, CGS utiliza la correlación como función objetivo en lugar del error medio. Esto minimiza naturalmente el peso de los valores atípicos.

6.5. Conclusiones

En este capítulo se utiliza un modelo CNN-CGS para abordar la evaluación estética del conjunto de datos Photo.net. Los resultados experimentales indican que el rendimiento global del modelo híbrido de extractor de características y sistema de predicción es superior a los métodos anteriores. Los resultados demuestran que las combinaciones de características de aprendizaje por transferencia de CNN y el método de aprendizaje automático híbrido pueden lograr resultados robustos y mejores que los métodos del estado del arte. Con un valor AUROC de 0,929 se logró el mejor resultado en comparación con el estado del arte.

Uno de los problemas de las métricas ad-hoc es el coste del tiempo de computación que necesitan algunas operaciones gráficas por ordenador. Debido al rápido procesamiento de las CNN, el uso de las características basadas en las CNN reduce la parte más costosa desde el punto de vista informático del método de clasificación y predicción.

Capítulo 7

Validación de las propuestas en una aplicación práctica

En este último capítulo de la parte experimental se validará, en un caso práctico real, tanto la capacidad de adaptación de la metodología propuesta en el capítulo 5, como la capacidad de aplicación comercial del sistema que obtuvo mejores resultados en el capítulo 6.

7.1. Contexto

En una realidad en la que un gran número de consumidores están conectados a Internet, son innumerables los estímulos y las opciones que reciben cada día. Sobre esta base, no es de extrañar que cada vez se elijan más productos en función del valor estético y el carácter distintivo del diseño visual [14]. Por otro lado, las empresas empiezan a ver la necesidad de encontrar nuevas formas de ser competitivas, de atraer y retener a los clientes y de mantener o aumentar las ventas [112, 113, 160]. Como afirman Deng y Poole [40], esto requiere un cambio de enfoque en el diseño y la experiencia del usuario y un mayor énfasis en la estética y las respuestas emocionales que suscita. Mediante la estimulación de respuestas emocionales, la estética puede contribuir al éxito del comercio electrónico y fomentar los comportamientos deseables del usuario, como pasar más tiempo navegando, explorar productos más variados, responder mejor a las iniciativas promocionales y aumentar la probabilidad de compra [40]. En el comercio electrónico, los productos se perciben a través de sus fotografías y es en ellas donde los consumidores juzgan el valor estético comentado. Además, tras la pandemia de COVID-19, es-

ta adaptación debe ser aun mayor en el ámbito digital, ya que ha contribuido al aumento de la digitalización de los consumidores [130].

Los grandes avances en las tecnologías relacionadas con la captura de imágenes y el crecimiento de otros campos como la informática visual [145, 175] han dado al usuario medio más poder y capacidad en la adquisición de imágenes. Desde el punto de vista del usuario, estas nuevas tecnologías crean la expectativa de imágenes más atractivas, pero esto también requiere el conocimiento y la ejecución de principios estéticos básicos durante la captura y la edición [5]. En este contexto, disponer de sistemas que proporcionen un criterio estético automático es de gran relevancia práctica.

El objetivo de la evaluación estética computacional es simular el sistema visual y la percepción humana para emitir un juicio estético sobre las imágenes de forma automática [167]. En los últimos años, muchos investigadores de diferentes áreas de conocimiento, como la Inteligencia Artificial, la Psicología, las Artes o el Diseño, se han centrado en la identificación de las características más relacionadas con las preferencias estéticas humanas, así como en el modelo de sistemas informáticos para recrear las evaluaciones humanas en tareas de clasificación y predicción [26, 64, 77, 83, 119, 138, 145, 146, 148, 171].

El modelo CGS, descrito en el capítulo 6 ha sido implementado para predecir el valor estético de imágenes pertenecientes al sector inmobiliario con el objetivo de mejorar el atractivo de los anuncios inmobiliarios. El precio de la vivienda y las características estéticas de la misma se entienden como los parámetros que más influyen en la decisión de compra de los individuos [6]. Sin una imagen que muestre de forma atractiva esas características estéticas, es difícil que el potencial comprador acceda al inmueble en una actualidad en la que la vía online es la más común en la búsqueda de vivienda. Además, es habitual que la decisión de visitar un inmueble que se ajusta a las características buscadas esté marcada por el atractivo de sus imágenes.

En este trabajo se utilizará el escaparate online de tres portes y tres agencias inmobiliarias diferentes y se estudiará la diferencia de impacto entre los anuncios que aparecen por defecto en sus webs y los anuncios propuestos por CGS en función de su valor estético.

7.2. Trabajos relacionados

En los últimos años han aumentado las publicaciones relacionadas con el estudio de las imágenes digitales y sus características [145]. Sin embargo, existen pocos trabajos de Estética Computacional o Inteligencia Artificial aplicada al arte y a la estética que contemplen aplicaciones comerciales o simplemente aplicaciones en el mundo real. El número de trabajos es aun menor si hablamos de presentar metodologías de validación de sistemas de Estética Computacional. A continuación, se presentan dos trabajos que plantean una propuesta de uso de la estética en el comercio electrónico, aunque centrados en un sector diferente al que plantea este trabajo.

Chen y Allebach [27] desarrollaron un sistema que puede predecir la calidad estética de las fotografías de productos de moda. Diseñaron un conjunto de características para la inferencia de la calidad estética en el contexto de las compras de moda online. También realizaron experimentos psicofísicos para construir una base de datos de 500 fotos calificadas manualmente, concretamente para las imágenes de un sitio web de compras de moda online. Más tarde, Wang y Allebach [164] introducen nuevas características de imagen y aplican la metodología de selección de características envolventes con el algoritmo de búsqueda “best-first”¹ para establecer un conjunto óptimo de características que produzcan la mejor precisión predictiva. Con ello, consiguen mejorar en un 19,3 % el resultado de Chen y Allebach [27], alcanzando un valor $RMSE = 1,687$.

También relacionado con el sector de la moda, pero con un enfoque más cercano al que se propone en este experimento, está el trabajo de Yu et al. [174]. En este caso, proponen introducir la información estética, muy relevante para la preferencia del usuario, en los sistemas de recomendación de prendas de vestir. Utilizaron el paquete de aprendizaje automático Weka [57] y entrenaron un modelo SVM utilizando 500 imágenes. Llevaron a cabo una 5-fold cross-validation para entrenar los parámetros del modelo y garantizar su fiabilidad. Consiguieron el mejor resultado con un valor de $RMSE = 2,09$ cuando utilizaron todos los conjuntos de características. Sus resultados demuestran que su enfoque puede captar la preferencia estética de los usuarios y superar significativamente los métodos de

¹El algoritmo *best-first* combina las ventajas de los algoritmos de búsqueda en profundidad y búsqueda en anchura. En este algoritmo se selecciona un nodo para la expansión basada en una función de evaluación. Esta función de evaluación devuelve un valor que sirve para representar lo deseable o indeseable que sería la expansión de un nodo. Se expande primero aquel nodo que tiene mejor evaluación.

recomendación más avanzados.

7.3. Aplicación comercial

El sistema de evaluación descrito en el capítulo 6 tiene un gran potencial en el comercio electrónico, ya que permite ordenar diferentes productos comerciales según el valor estético de sus imágenes.

La aplicación de este sistema en los escaparates electrónicos permite que los productos con las imágenes más atractivas se muestren en las primeras posiciones de la web. De forma directa, esto debería suponer un aumento en el número de clics que los visitantes realizan en la web, ya que encontrarían primero los productos que más les interesan. Esta premisa será validada con la parte experimental de esta propuesta.

Indirectamente, si confirmamos que la premisa anterior es cierta, el sistema de ordenación según el atractivo comercial de las imágenes también consigue (i) mejorar la experiencia del usuario, (ii) mejorar el posicionamiento de la empresa y (iii) aumentar la captación de nuevos clientes.

7.4. Materiales y métodos

En este trabajo, la evaluación estética de las fotografías se realiza a través del algoritmo genético híbrido que ofreció mejores resultados en el capítulo 6: CGS. Además, se creó un dataset específico para el entrenamiento del sistema y otro para su validación práctica.

7.4.1. Dataset de entrenamiento

El sistema descrito anteriormente fue entrenado con un dataset formado por imágenes del parque inmobiliario español, con más de un millón de valoraciones humanas, enfocadas a tareas de incremento del impacto comercial en el sector inmobiliario.

Para crear el conjunto de datos, se utilizó la metodología diseñada y presentada en el capítulo 5. Se seleccionaron aleatoriamente miles de anuncios inmobiliarios existentes en los portales inmobiliarios españoles y se llevó a cabo un proceso de control de la calidad estética de las imágenes de forma que se clasificaron

según su valor estético en “alto”, “medio” y “bajo”. Como es habitual, y como ya veíamos en la distribución de las imágenes del dataset creado en el capítulo 5, existía un mayor número de imágenes que tenían un valor estético medio, por lo que se nutrió el dataset con imágenes buscadas de forma manual que tuviesen un valor estético más bajo que la media y también más alto. De esta forma se obtuvo un conjunto representativo para el entrenamiento.

Cada una de las imágenes del conjunto creado fue evaluada por un grupo de humanos mediante la plataforma online Amazon Mechanical Turk (AMT).

AMT es un mercado de *crowdsourcing* que facilita a los individuos y a las empresas la subcontratación de sus procesos y trabajos a una mano de obra distribuida que puede realizar estas tareas virtualmente. Esto puede incluir desde la validación de datos de investigación hasta tareas más subjetivas como la participación en encuestas, la moderación de contenidos, etc. AMT permite crear una tarea, elegir el perfil y el número de trabajadores que deben realizar dicha tarea y estipular la cantidad que se pagará a cada trabajador por ello. Este recurso online permite ahorrar costes en la fase de construcción de los datasets, ya que no es necesaria una fase de reclutamiento de los trabajadores ni contar con un lugar físico para la realización del proceso de forma presencial. Sin embargo, es necesario ser prudente a la hora de utilizar AMT, ya que no se puede tener tanto control sobre la fase de evaluación como en un test presencial, por lo que se deben realizar pruebas de control de las tareas para asegurarse, por ejemplo, de que no se recogen respuestas aleatorias. Por otra parte, cuando se indica el número de trabajadores que deben realizar la tarea, AMT se asegura de que cada tarea tenga ese número de evaluaciones pero no de que todas pertenezcan a los mismos trabajadores, es decir, un trabajador puede dejar la tarea a la mitad y será otro el que la finalice para alcanzar el número estipulado. Esto último no debería ser un problema al trabajar con medias globales, si se define bien el perfil de la población.

En este caso concreto, la tarea de los evaluadores en AMT fue determinar el atractivo comercial de cada imagen puntuándola en una escala de entre 1 y 10, donde 1 corresponde al atractivo más bajo y 10, al más alto. Cada imagen fue evaluada por un mínimo de 20 personas.

Al tratarse de un proceso de evaluación online, el control fue menor que en un experimento presencial, por lo que se llevaron a cabo otros procesos de control para evitar evaluaciones aleatorias de los participantes. En concreto, se valoró el

tiempo de respuesta de cada participante por imagen y la correlación de cada uno de ellos con la media. Si la correlación de sus respuestas era inferior a un valor de 0,5, se rechazaban las valoraciones de ese participante.

Después de estos procesos de control, cada imagen se asoció con su puntuación media que corresponde con el atractivo comercial medio en España (todos los evaluadores son españoles, igual que todos los inmuebles de los que se extrajeron las imágenes). Una imagen con una puntuación alta probablemente tendrá un alto número de visualizaciones si se utiliza como imagen principal del anuncio inmobiliario.

7.4.2. Dataset de evaluación

El objetivo básico de este experimento es comprobar cuántos clics reciben los inmuebles que aparecen por defecto en las webs inmobiliarias y cuántos clics recibirían si esos inmuebles fueran los de mayor valor estético. Para ello, se ha creado un nuevo conjunto de datos con imágenes de diferentes anuncios inmobiliarios recogidos de las webs de tres portales y tres agencias inmobiliarias de España. Las fuentes han sido anonimizadas, por lo que a partir de ahora se tratarán como Portal1, Portal2, Portal3 y Agencia1, Agencia2, Agencia3.

El método de composición de los conjuntos de imágenes se presenta en el diagrama 7.1 y se describe de forma detallada a continuación:

1. En primer lugar, se realiza una búsqueda de inmuebles en cada uno de los sitios web. En el caso de los portales inmobiliarios, la localización se limita a la ciudad de A Coruña. En el caso de las agencias inmobiliarias, no se aplica ningún filtro de búsqueda, ya que son empresas que cubren un territorio más reducido.
2. Después, se selecciona un conjunto representativo de los inmuebles, que estará formado por los primeros resultados de la búsqueda (paso 1 en el diagrama 7.1). En el caso de los portales serán los 250 primeros y, en el caso de las agencias inmobiliarias, serán los 125 primeros. Esta diferencia se debe únicamente al volumen de inmuebles que manejan las propias plataformas. De este conjunto se selecciona el 20% de los inmuebles que aparecen en primer lugar en la búsqueda (paso 2.1 en el diagrama 7.1), que formarán el conjunto denominado “original” de cada fuente. En el caso de los portales,

serán los 50 primeros inmuebles que aparezcan en la búsqueda y, en el caso de las agencias, serán los 25 primeros.

- El siguiente paso es formar el conjunto denominado “estético” de cada fuente. Para ello, se evalúa la primera imagen de cada anuncio del conjunto representativo inicial (250 en caso de los portales y 125 en el de las agencias), mediante el sistema de evaluación CGS (paso 2.2 en el diagrama 7.1). Una vez evaluadas todas las imágenes, se selecciona el 20 % con mayor puntuación y con ellas se constituye el conjunto “estético” (paso 3 en el diagrama 7.1).

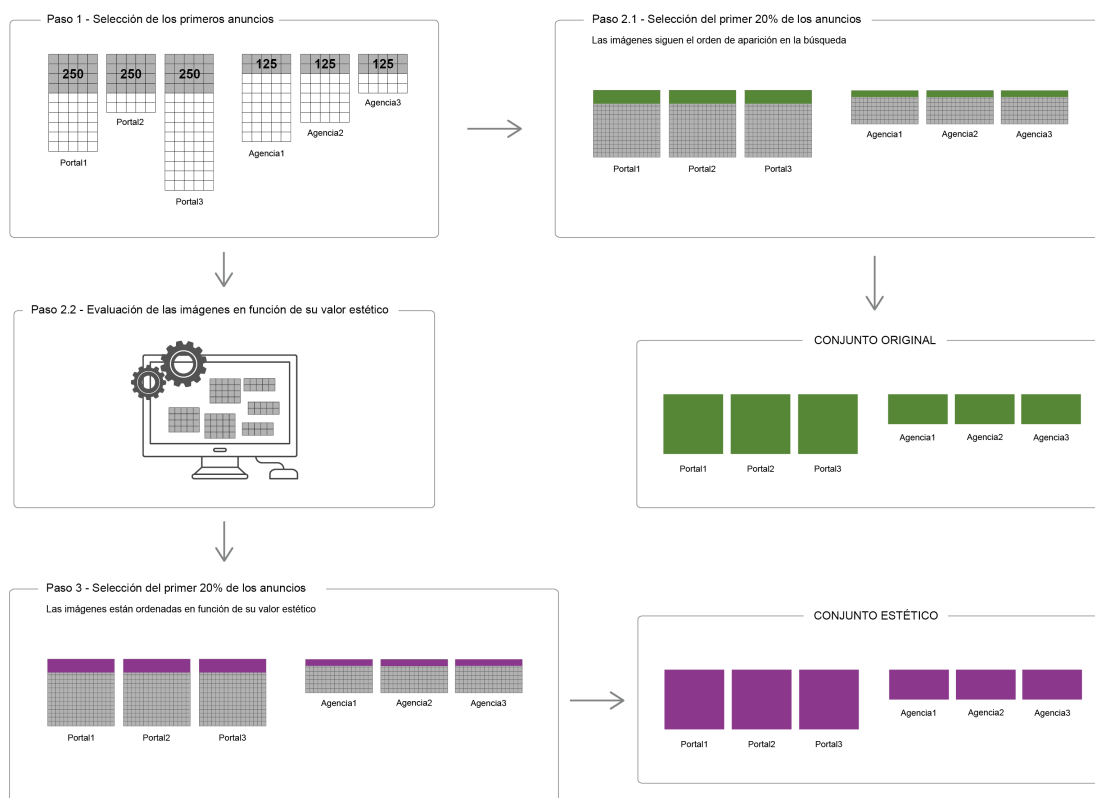
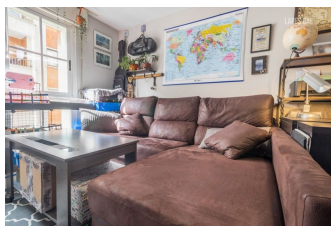


Figura 7.1: Diagrama que muestra el proceso de creación del dataset de evaluación.

De este modo, tenemos 12 conjuntos diferentes en total, 2 para cada caso de estudio (para cada portal y para cada agencia inmobiliaria): el conjunto original con el 20 % de las imágenes que aparecen primero en la búsqueda y el conjunto estético con el 20 % de las imágenes mejor evaluadas por el sistema de valoración

estética. Así, cada portal tiene un conjunto original de 50 imágenes y un conjunto estético de 50 imágenes; y cada agencia tiene un conjunto original de 25 imágenes y un conjunto estético de 25 imágenes.

En las figuras 7.2 y 7.3 se presenta una muestra de los dos conjuntos de la Agencia1. La figura 7.2 muestra las nueve imágenes que aparecen por defecto en primer lugar en la búsqueda y la figura 7.3 muestra las nueve con mayor valoración estética. Sin entrar en detalle, se puede observar una diferencia importante entre los dos conjuntos: las imágenes del conjunto original pertenecen, en su mayoría, a habitaciones del interior de las casas y la mayor parte del conjunto estético son fotografías del exterior.



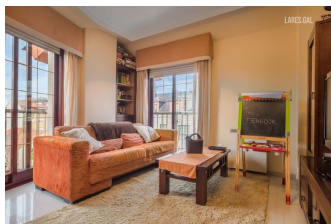
(a) Primera



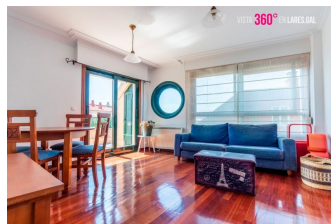
(b) Segunda



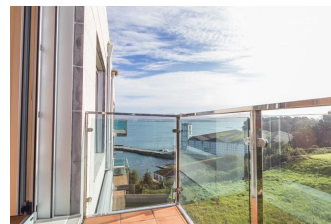
(c) Tercera



(d) Cuarta



(e) Quinta



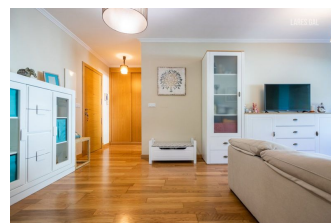
(f) Sexta



(g) Séptima



(h) Octava



(i) Novena

Figura 7.2: Imágenes de los nueve primeros anuncios del conjunto original de una de la Agencia1.

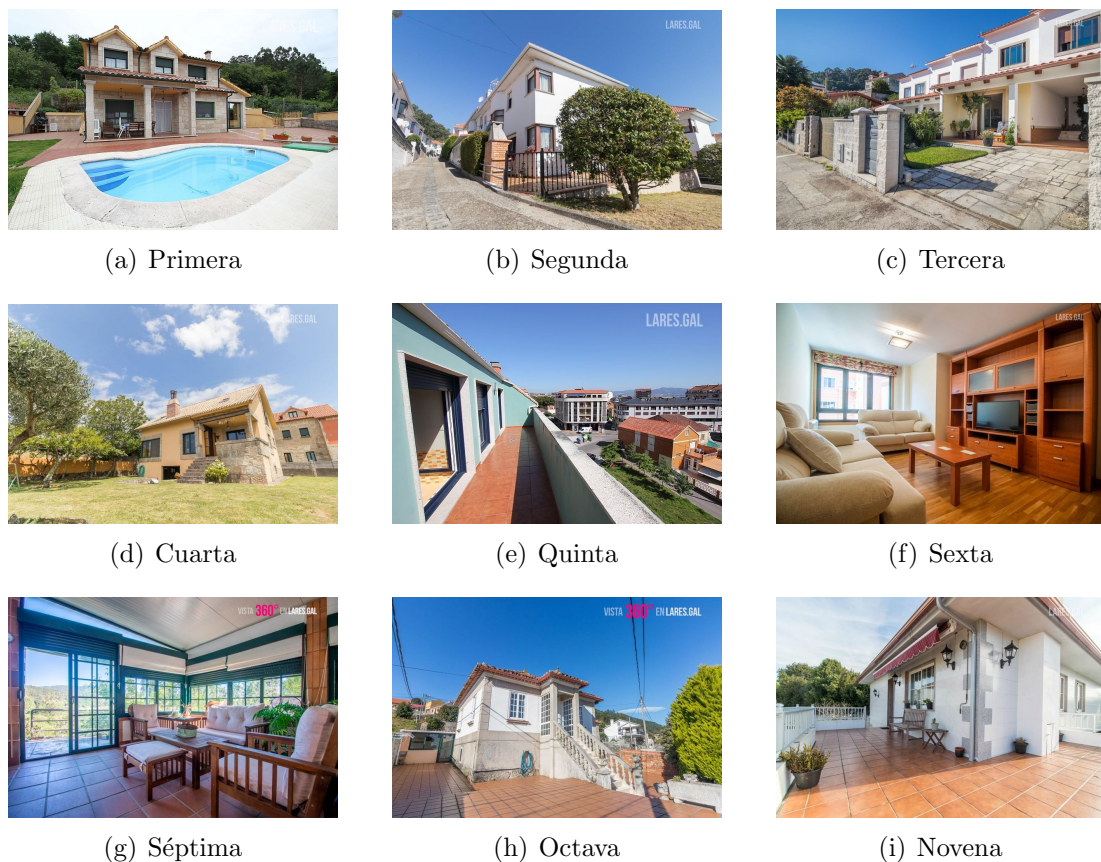


Figura 7.3: Imágenes del conjunto estético de la Agencia1 propuestas para las primeras nueve posiciones.

Al formar los conjuntos de imágenes con el método descrito anteriormente, existe la posibilidad de que algunas imágenes estén presentes en ambos conjuntos, tanto en el original como en el estético. Por ejemplo, puede darse el caso de que la imagen con mayor valor estético de una de las agencias esté entre los primeros 25 resultados de la búsqueda. La tabla 7.1 muestra el número de imágenes repetidas en cada caso de estudio y el porcentaje que representan del total de imágenes de cada fuente. En este caso, hemos decidido mantener las imágenes repetidas en el experimento y contar el número de votos que recibieron para el resultado de ambos conjuntos.

7.4.3. Muestra y procedimiento

Como se ha comentado anteriormente, este experimento pretende comparar el número de clics que recibirían los inmuebles que aparecen por defecto en la

Cuadro 7.1: Número de imágenes repetidas en el conjunto original y en el conjunto estético por cada caso de estudio.

Conjunto	Imágenes repetidas	Proporción de la muestra (%)
Portal1	6	12
Portal2	10	20
Portal3	9	18
Agencia1	24	48
Agencia2	16	32
Agencia3	16	32

búsqueda de cada portal o agencia inmobiliaria y el número de clics que recibirían los propuestos por el sistema evolutivo de evaluación estética. Para ello, se realizó una encuesta a un grupo de personas a través de la herramienta Amazon Mechanical Turk (AMT).

En este experimento, el proceso metodológico diseñado tiene gran importancia, ya que ciertos detalles podrían sesgar la respuesta de los votantes.

En primer lugar, las imágenes de los 12 conjuntos descritos en el apartado 7.4.2 se mezclan de forma aleatoria, sin tener en cuenta el conjunto al que pertenecen ni la fuente de origen. A continuación, se dividen en grupos de 10 imágenes, también aleatorios, de modo que en cada grupo puede haber imágenes de diferentes conjuntos y de diferentes fuentes. Con estos grupos se diseña la encuesta de AMT.

La encuesta consiste en un enunciado en el que se explica a los votantes el objetivo de la encuesta y, a continuación, se muestran 10 imágenes (solo se muestra un grupo de cada vez), cada una de ellas con dos opciones al lado: una que indica que sí haría clic en esa imagen y la otra que indica lo contrario. La figura 7.4 muestra la encuesta tal y como la vería el o la participante. Los evaluadores deben marcar una opción u otra, en función de si harían clic en esa imagen si la vieran en una web inmobiliaria o no. No hay un número determinado de imágenes que marcar de forma positiva, sino que cada votante puede indicar el número de imágenes que consideraría en una situación real. Por defecto, la respuesta marcada en cada imagen es “no”, por lo que cada votante debe indicar explícitamente lo contrario si cree que es susceptible de recibir un clic.

A la hora de diseñar la encuesta se definió que cada imagen debía ser evaluada por 50 personas. Sin embargo, como se comentaba anteriormente, AMT no garantiza que todas las imágenes sean evaluadas por las mismas 50 personas.

Así, un total de 61 personas participaron en la encuesta, aunque cada imagen fue evaluada por solo 50 personas, como se indicó. Esto se debe a que AMT permite abandonar la encuesta en cualquier momento, por lo que si una persona evalúa solo el 30 % de las imágenes, el 70 % restante se mostrará a otro participante que no haya pasado por ellas.

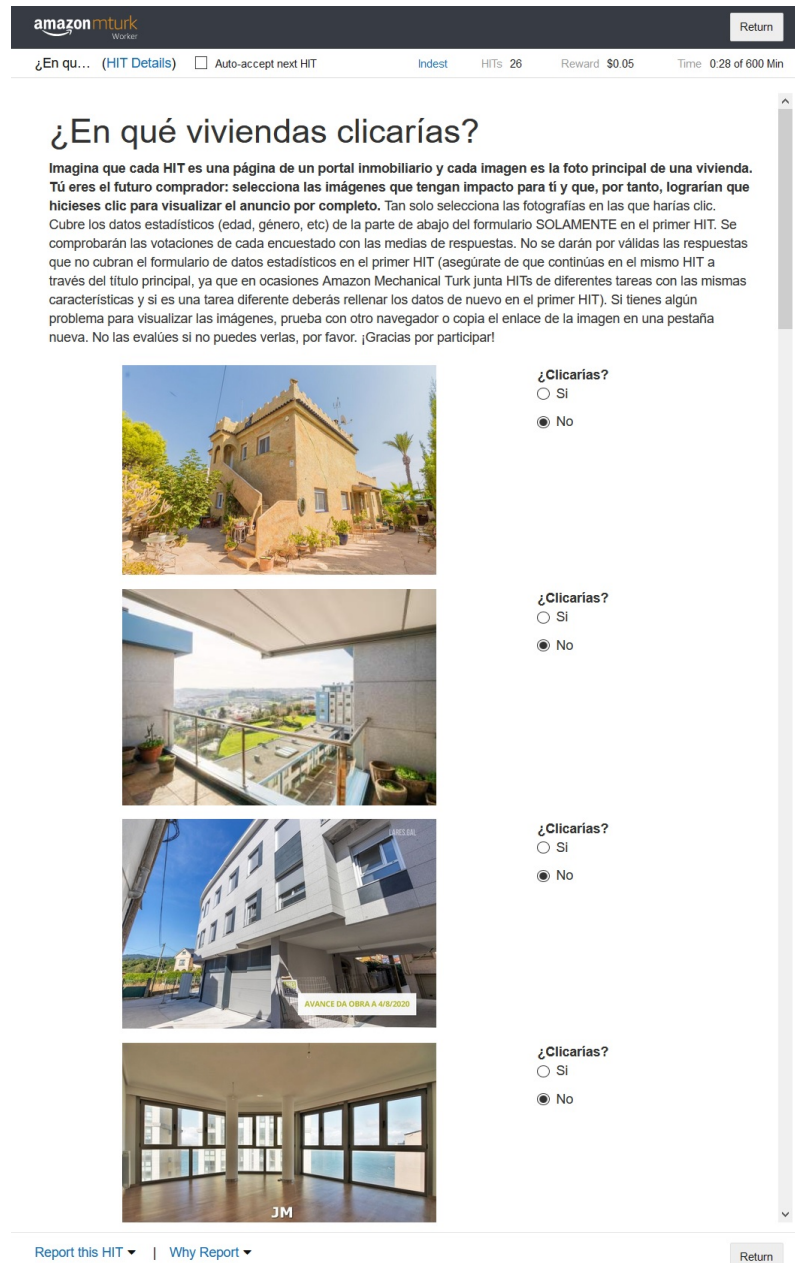


Figura 7.4: Ejemplo de la encuesta desde la perspectiva de los participantes.

Cuadro 7.2: Datos personales proporcionados por los participantes de la encuesta.

Elemento	Frecuencia	Proporción de la muestra (%)
Género		
Mujer	15	30,6
Hombre	34	69,4
Edad		
25 e inferior	5	10,2
26–35	18	36,7
36–45	21	42,9
46 y superior	5	10,2
Ingresos		
600€ e inferior	4	8,2
601€–1200€	13	26,5
1201€–1600€	11	22,5
1601€–2000€	10	20,4
2001€–2500€	3	6,1
2501€–3000€	3	6,1
3001€ y superior	5	10,2
Actividades relacionadas		
Sí	5	10,4
No	43	89,6
Habitat		
Urbano	37	75,5
Semi-urbano	9	18,4
Rural	3	6,1
Convivencia		
Solo	5	10,2
Con pareja sin hijos	16	32,7
Con pareja e hijos	17	34,7
Con padres u otros parientes	8	16,3
Con otra gente (no parientes)	3	6,1
Propiedad		
Alquiler	21	42,9
En propiedad	25	51
Otros	3	6,1
Tipo de vivienda		
Casa	9	18,4
Piso/Apartamento	40	81,6

La encuesta tiene una última parte en la que se solicitan algunos datos personales relacionados con el ámbito de la tarea. En este caso, solo el 81,6% de

los participantes respondió a esta parte de la encuesta, aunque es un porcentaje significativo y aporta datos relevantes para el estudio. En la tabla 7.2 se pueden ver los resultados obtenidos.

El grupo de participantes de este experimento está compuesto por españoles de entre 18 y 48 años con una edad media de 35,6 años. Estos datos son relevantes en el estudio porque las imágenes que se evaluaron pertenecen al panorama inmobiliario español, por lo que los participantes deben ser potenciales demandantes de vivienda en España. Además, la mayoría de ellos (89,6%) no tienen estudios ni actividad profesional relacionada con la fotografía, el diseño, la arquitectura o actividades similares que puedan sesgar su percepción de las imágenes. Por otro lado, el 93,9% vive en una vivienda en propiedad o en alquiler, por lo que han sido y pueden seguir siendo potenciales buscadores de vivienda.

7.5. Resultados

Esta sección presenta los resultados obtenidos en la encuesta realizada en AMT, con 50 evaluaciones por imagen. Los resultados se dividieron en dos tablas, según el origen de las imágenes del conjunto. La tabla 7.3 muestra los resultados para cada uno de los tres portales inmobiliarios que se tomaron para el estudio. La tabla 7.4 muestra los resultados obtenidos para las imágenes de las agencias inmobiliarias.

Cuadro 7.3: Número medio de votos recibidos por imagen para cada conjunto, el incremento por cada portal y el incremento medio total en los tres portales. El número máximo de votos por imagen es 50.

	Portal1	Portal2	Portal3	Media
Original	21,40	17,45	16,02	18,29
Estético	28,40	26,80	28,52	27,90
Incremento	32,71 %	53,58 %	78,03 %	52,54 %

Cuadro 7.4: Número medio de votos recibidos por imagen para cada conjunto, el incremento para cada agencia y el incremento medio total en las tres agencias inmobiliarias. El número máximo de votos por imagen es 50.

	Agencia1	Agencia2	Agencia3	Media
Original	21,60	30,21	11,08	20,96
Estético	28,76	37,24	22,08	29,36
Incremento	33,15 %	23,27 %	99,28 %	40,08 %

El resultado más destacado de este experimento ha sido el hecho de que todos los conjuntos estéticos superan en clics a los conjuntos originales, tanto en los portales como en las agencias inmobiliarias.

En el caso de los portales, como se muestra en la tabla 7.3, los conjuntos estéticos consiguen un incremento total medio del 52,54 %. El menor incremento se registra en el Portal1, con un 32,71 % y el mayor, en el Portal3, con un 78,03 % pasando por el Portal2, con un 53,58 %. También hay que destacar que la media de votos por imagen en los conjuntos estéticos de los tres portales es muy similar, lo que sugiere que los tres portales tienen imágenes de similar valor estético, solo que por defecto se ordenan con un criterio diferente.

Por otro lado, las agencias inmobiliarias revelan resultados mucho más variados, como muestra la tabla 7.4. En este caso, los conjuntos estéticos alcanzan un incremento total medio del 40,08 %. El mayor y más destacado incremento en este grupo es el de la Agencia3, con un 99,28 %, seguido del 33,12 % de la Agencia1 y del 23,27 % de la Agencia2. La media de votos por imagen en cada uno de los grupos estéticos es también más variada, con 37,24 para la Agencia2, a pesar de ser la agencia con menor incremento en el número de clics. Esto se debe probablemente a que las imágenes de esta agencia tienen un mayor valor estético. En el caso de la Agencia3 se deduce que, además de tener fotografías de menor valor estético (tiene el menor número medio de votos por imagen), el orden que aparece por defecto en la búsqueda no es atractivo para los usuarios.

Las figuras 7.5 y 7.6 presentan una muestra de las imágenes con más votos afirmativos en ambos conjuntos de la Agencia1. Como se puede observar y como ya se mostró en la tabla de resultados, las imágenes del conjunto estético han recibido más votos que las del conjunto original, siendo las más votadas superadas por 10 votos.



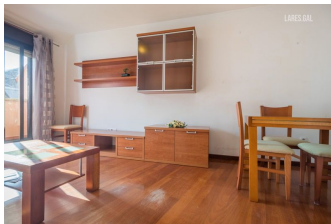
(a) 35 votos



(b) 32 votos



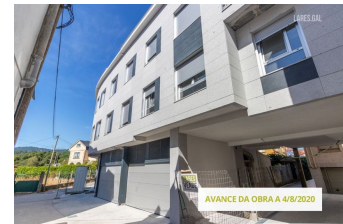
(c) 28 votos



(d) 23 votos



(e) 20 votos



(f) 20 votos



(g) 19 votos



(h) 18 votos



(i) 17 votos

Figura 7.5: Imágenes con más votos afirmativos del conjunto original de la Agencia1.

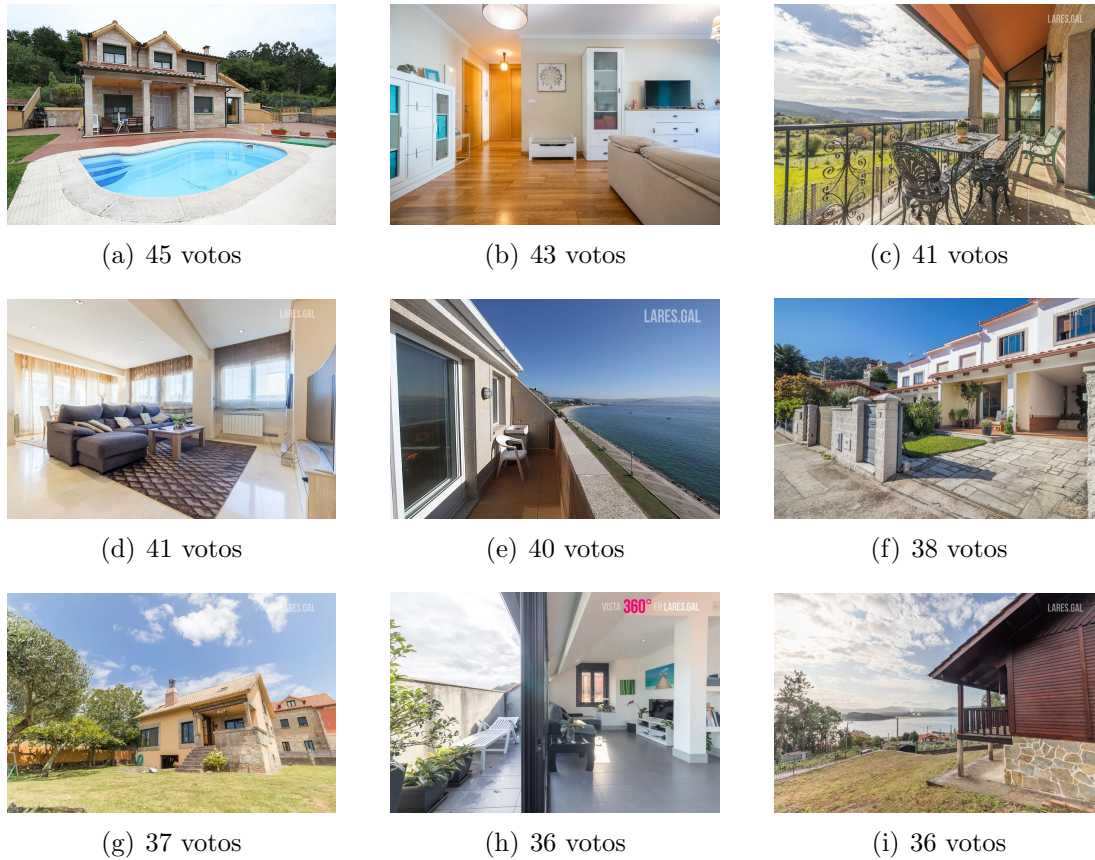


Figura 7.6: Imágenes con más votos afirmativos del conjunto estético de la Agencia1.

En el apéndice C se puede encontrar una muestra comparativa de las imágenes más extensa, donde se presentan según su orden original, su orden estético y el número de votos recibidos.

7.5.1. Análisis estadístico

Con los resultados obtenidos se realizó un análisis estadístico para comprobar si la diferencia de clics entre el conjunto de imágenes original y el conjunto estético era significativa.

En primer lugar, y siguiendo a Guyon et al. [67], se realizó la prueba de bondad de ajuste de Kolmogorov-Smirnov (K-S) a cada uno de los conjuntos para comprobar si los datos seguían una distribución normal. La tabla 7.5 muestra los datos estadísticos de cada uno de los conjuntos. Esta información permite decidir si se realiza una prueba paramétrica o una prueba no paramétrica para comprobar si las diferencias entre los conjuntos son significativas.

Cuadro 7.5: Datos estadísticos para cada conjunto: el conjunto de imágenes original y el estético.

Valor	Original	Estético
Cuenta	225	225
Media	19,15771812	27,37288136
Mediana	17	29
Desviación estándar	12,28390138	11,44299433
Asimetría	0,328208	-0,346432
Curtosis	-1,118261	-0,943309
D de Kolmogorov	0,10467	0,10211
<i>p</i> -value	0,00292	0,004259

La hipótesis nula a contrastar con esta prueba K-S es que los datos analizados siguen una distribución normal. El valor D de Kolmogorov es la mayor diferencia absoluta entre la frecuencia acumulada observada y la frecuencia acumulada teórica, obtenida a partir de una distribución normal. Cuanto mayor sea la discrepancia entre la distribución empírica y la distribución teórica, mayor será el valor de D . El criterio para aceptar la hipótesis nula es que D debe ser menor o igual que $D\alpha$. En este caso, al tener una muestra de 225 datos y estar comprobando una distribución normal, el valor $D\alpha$ es de 0,059489. El valor D de Kolmogorov es 0,10467 para las imágenes originales y 0,10211 para las imágenes del conjunto estético. Además, el p -value es de 0,00292 y 0,004259 respectivamente. Por tanto, al ser $D > D\alpha$ y contar con un p -value $< 0,05$, se puede concluir que ninguno de los conjuntos sigue una distribución normal.

Al observar que los datos no siguen una distribución normal, se decide realizar una prueba no paramétrica, concretamente la de Wilcoxon [43, 120, 169]. La elección de esta prueba se basa en el estado del arte de la metodología, donde se recomienda su uso en los casos en que la comparación se realiza solo entre dos muestras con alguna relación entre los participantes [7, 111, 162].

Para esta prueba planteamos una hipótesis nula (H_0) y una alternativa (H_1):
 (H_0) : no hay diferencia entre el número de clics que reciben las imágenes originales y el que reciben las imágenes del conjunto estético.

(H_1) : sí existe diferencia entre el número de clics que reciben las imágenes originales y el que reciben las imágenes del conjunto estético.

El valor obtenido ha sido $W = 27,392$, con un p -value de $1,978 e^{-15}$. Este p -value permite rechazar la hipótesis nula, por lo que se puede concluir que la

diferencia de clics entre ambos conjuntos sí es significativa.

Podemos ver estos resultados de forma gráfica en la figura 7.7. En este gráfico se observa el porcentaje de imágenes para cada número de votos hasta un máximo de 50. Las barras de color verde claro muestran el porcentaje de clics de las imágenes seleccionadas en los anuncios originales por los propios agentes inmobiliarios, mientras que las barras de color verde oscuro muestran el porcentaje de clics obtenidos en las imágenes propuestas por el sistema de IA. Así, se puede observar una tendencia descendente de las imágenes del conjunto original a medida que aumenta la cuenta de clics y, por el contrario, una tendencia ascendente de las imágenes del conjunto estético.

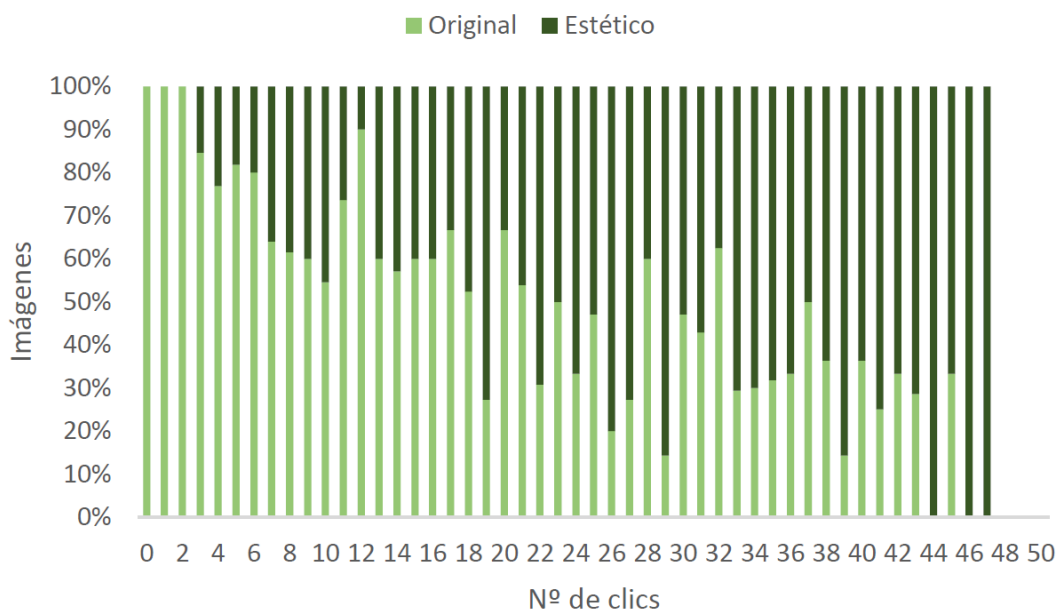


Figura 7.7: Porcentaje de imágenes de cada conjunto por cada número de clics. El eje x define el número de clics con un máximo de 50 y el eje y representa el porcentaje de imágenes de cada conjunto que recibieron esos clics. Verde claro: imágenes de los conjuntos originales. Verde oscuro: imágenes de los conjuntos estéticos.

7.6. Discusión

Los resultados del experimento fueron claros: las páginas web de los portales e inmobiliarias recibirían más clics si tuvieran en cuenta el valor estético de sus

imágenes a la hora de ofrecer los resultados de la búsqueda. De esta forma, se obtendrían respuestas emocionales estimuladas a las que se refieren Deng y Poole [40] y se convertirían en más tiempo de navegación por la web, explorando más propiedades o aumentando la probabilidad de visitar el inmueble.

A partir del análisis estadístico realizado, se ha podido identificar, como se puede ver en la figura 7.8, que la selección de imágenes basada en el valor estético tiene una influencia significativa en el número de clics. Si observamos los datos del conjunto original (verde), podemos ver que el 50 % de las imágenes tienen menos de 17 clics. Tiene una asimetría positiva, pero con un alto rango de dispersión en su distribución. En cuanto al conjunto estético (violeta), la mediana es de 29 clics y, aunque tiene un sesgo negativo, su rango de dispersión es menor en este conjunto de datos. El p -value obtenido con Wilcoxon nos permite afirmar que la diferencia entre el número de clics obtenidos en el conjunto original y el número de clics obtenidos en el conjunto estético son estadísticamente significativas. Por tanto, se puede observar que se obtienen mejores resultados con el uso del conjunto de datos estético.

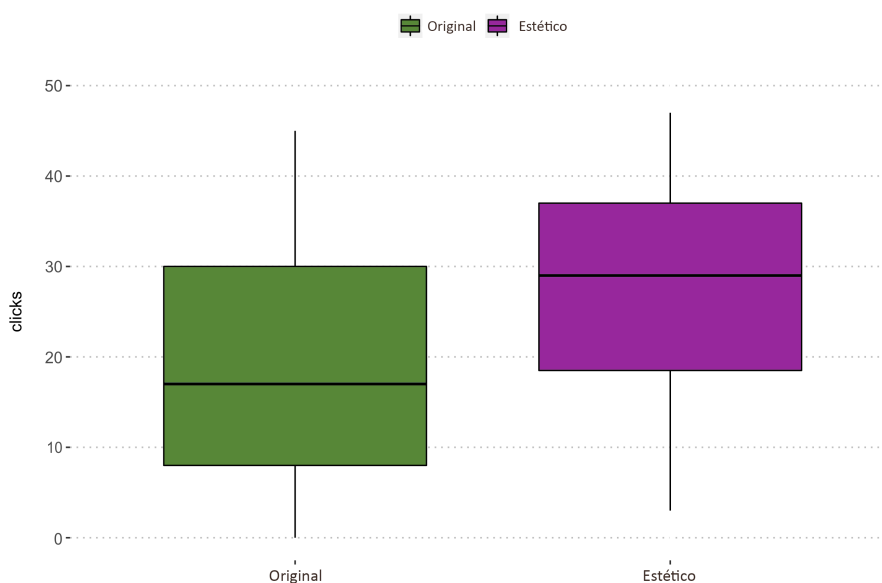


Figura 7.8: Boxplot en el que se muestra la distribución de ambos conjuntos de datos. Verde: clics del conjunto original. Violeta: clics del conjunto estético.

Merece una mención especial la coincidencia entre la imagen más votada del conjunto estético de la Agencia1 (ver (a) de la figura 7.6) y la imagen propuesta para la primera posición de esa agencia por el sistema de IA (ver (a) de la figura

7.3). El sistema propuso para la primera posición la imagen que recibiría más clics de los potenciales compradores.

Los portales y las agencias inmobiliarias utilizan diferentes sistemas para ordenar los inmuebles por defecto en los resultados de búsqueda. Algunos de ellos colocan en las primeras posiciones los anuncios más recientes, mientras que otros utilizan un criterio de relevancia que no está definido pero que puede tener en cuenta la existencia de anuncios promocionados (anuncios por los que se paga un dinero extra para ocupar posiciones destacadas).

Algo que se puede deducir de los resultados obtenidos es la calidad estética de las fotografías de los portales y agencias (cuánto cuidan la estética de sus imágenes) por el número de votos medios recibidos. En el caso de los portales, el Portal1 cuenta con una media de 21,4 clics, frente al Portal3 que cuenta con una media de 16,02. Sin embargo, los conjuntos estéticos de los tres portales cuentan con valores medios similares. Esto sugiere que la calidad estética de sus fotografías es semejante, pero utilizan criterios diferentes para su ordenación. En el caso de las agencias, la Agencia1 parece tener niveles estéticos similares a los de Portal1, pero las grandes diferencias aparecen en Agencia2 y Agencia3. La media de votos recibidos por estas dos sugiere que la Agencia2 tiene fotografías y anuncios de mayor valor estético que el resto de las agencias y portales, mientras que la Agencia3 presenta imágenes con un valor estético notablemente inferior.

En la sección 7.4.2 se ha explicado cómo se seleccionó el 20 % del conjunto representativo para formar los dos conjuntos de cada fuente. Sin embargo, también se ha analizado cómo cambian los resultados cuando varía el porcentaje de imágenes seleccionadas del conjunto total. La tabla 7.6 muestra el número medio de votos de las tres agencias inmobiliarias estudiadas, según el porcentaje de imágenes seleccionadas. En este caso, además del 20 %, hemos analizado la variación con el 10 % y el 40 %. Como se puede observar, en los conjuntos originales hay variaciones en la media de votos al estudiar un porcentaje diferente de imágenes. Esto es especialmente evidente en Agencia1, donde hay un aumento significativo de los votos medios si estudiamos solo el primer 10 % del conjunto. La Agencia3 también muestra mejores imágenes en el 10 % en comparación con el 40 %. Sin embargo, la Agencia2 no muestra ninguna variación significativa entre los distintos porcentajes.

Cuadro 7.6: Media de votos por imagen recibido por cada conjunto de las agencias según el porcentaje de imágenes analizadas.

	%	Agencia1	Agencia2	Agencia3	Media
Original	10	24,00	29,27	13,16	22,14
Original	20	21,60	30,21	11,08	20,96
Original	40	19,97	29,97	10,42	20,21
Estético	10	28,90	38,08	27,66	31,55
Estético	20	28,76	37,24	22,08	29,36
Estético	40	25,76	35,88	18,88	26,84

El hecho de que algunas agencias muestren una mejora en el número medio de clics cuando se estudia un porcentaje menor de la muestra, hace patente que existe un esfuerzo por parte de estas agencias para mostrar propiedades atractivas en las primeras posiciones. Se puede suponer que este esfuerzo se debe a que esas agencias consideran que mejora la imagen de la agencia y maximiza el número de clics. Sin embargo, los resultados obtenidos con el sistema informático son significativamente mejores que los de las agencias inmobiliarias.

7.7. Conclusiones

En este capítulo se demuestra la utilidad real de un sistema de evaluación estética en el marco de una aplicación comercial. En concreto, se ha demostrado un aumento significativo del impacto de los anuncios inmobiliarios al presentar imágenes con mayor valor estético.

Para ello, se han tomado como casos de estudio tres portales inmobiliarios y tres agencias inmobiliarias diferentes de España y se ha creado un conjunto de datos con imágenes recogidas de sus páginas web. A partir de ahí, se ha presentado un método de validación que se puede adaptar a diferentes problemas de estética computacional.

Se han obtenido resultados significativos, proponiendo un incremento medio en el número de clics en los anuncios del 52,54 % en el caso de los portales y del 40,08 % en el caso de las agencias inmobiliarias.

Existen también otras formas posibles de emplear el sistema de evaluación estética. Por ejemplo, se puede utilizar para seleccionar la mejor imagen de cada anuncio de modo que este reciba más visitas. Además, el sistema puede ser utilizado en otros dominios (coches, anuncios de segunda mano, comercio electrónico

en general...) simplemente creando un conjunto de datos evaluados por humanos en ese dominio.

Parte III

Conclusiones

Capítulo 8

Conclusiones y trabajo futuro

En este capítulo se exponen las conclusiones generales obtenidas en esta tesis doctoral y se presenta el trabajo que se pretende desarrollar en relación a esta línea de investigación en un futuro próximo.

8.1. Conclusiones

Esta tesis aborda diferentes problemas recurrentes en el campo de la estética computacional y propone nuevas soluciones que son validadas finalmente en un caso práctico real. En primer lugar, esta tesis estudia las limitaciones de los datasets utilizados en estética computacional y propone una nueva metodología para la creación de conjuntos de imágenes que se pueda aplicar en problemas de Machine Learning. Después, se expone un nuevo algoritmo genético híbrido para predicción estética desarrollado previamente en el grupo Redes de Neuronas Artificiales y Sistemas Adaptativos (RNASA) de la Universidade da Coruña, por Carballal [136]. Se utilizó *transfer learning* con este nuevo algoritmo genético y se compararon sus resultados con otros modelos del estado del arte. Finalmente, se aplicó a un caso práctico real tanto la metodología de creación de datasets propuesta en primer lugar, como el modelo híbrido que ofreció mejores resultados en la fase experimental. Los resultados sugieren que utilizar estas herramientas en la vida cotidiana pueden mejorar tanto la experiencia de los usuarios como el rendimiento de los *e-commerce*.

De forma más detallada, las conclusiones que se pueden extraer de esta tesis doctoral son:

- Tras detectar algunas limitaciones en los datasets utilizados en el estado

del arte en estética computacional, se propuso una nueva metodología para la construcción de conjuntos de datos con el objetivo de mejorar la calidad de entrenamiento de los modelos. Hay dos aspectos fundamentales en la propuesta de esta metodología en comparación con los datasets del estado del arte: en primer lugar, no se hace una clasificación binaria de las imágenes, sino que se distribuyen por rangos en una puntuación de entre 0 y 10, y cada rango cuenta con el mismo número de imágenes, en contraposición con la mayor parte de los datasets existentes que siguen una distribución normal; y, en segundo lugar, las imágenes son evaluadas una segunda vez por un grupo de humanos en condiciones experimentales controladas.

- Siguiendo la metodología propuesta, también se presenta un nuevo dataset formado por 1.000 imágenes del portal fotográfico DPChallenge, que cuenta con tres evaluaciones diferentes: (i) la puntuación media obtenida en DPChallenge, con al menos 100 valoraciones por imagen; (ii) la evaluación del valor estético realizada en condiciones experimentales controladas con un mínimo de 10 votos por imagen; y (iii) la evaluación de la calidad de las imágenes realizada en las mismas condiciones que la evaluación estética.
- Para probar la eficacia de las propuestas anteriores, se utilizaron diferentes métodos computacionales de última generación y se entrenaron para predecir los tres criterios (DPChallenge, estética y calidad), con lo que se obtuvieron correlaciones de 0,578, 0,456 y 0,539 respectivamente, utilizando como medida de rendimiento la ρ de Spearman, y de 0,574, 0,451 y 0,562, utilizando la r de Pearson. Estos resultados sugieren que las técnicas de aprendizaje automático tienen una mayor capacidad de aprender la percepción humana de la calidad técnica que otros aspectos más subjetivos como la estética.
- Una vez abordada la cuestión del conjunto de entrenamiento, se propuso un nuevo modelo para la medición estética basado en Machine Learning, denominado Correlation by Genetic Search (CGS). Los resultados demuestran que las combinaciones de características de aprendizaje por transferencia de CNN y el método de aprendizaje automático híbrido pueden lograr resultados robustos y estadísticamente mejores que los del estado del arte. Utilizando este nuevo método con la red GoogleNet se alcanzó un valor AUROC = 0,929 para la tarea de predicción estética, que supera el valor

AUROC = 0,896 del estado del arte.

- Se compararon los resultados obtenidos con CGS con los mejores resultados del estado del arte y se comprobó que el enfoque propuesto superaba al estado del arte tanto en clasificación como en predicción. Una de las razones por las que sucede esto puede ser que, al utilizar un conjunto de métricas no diseñadas por humanos, sea posible captar características de la imagen en las que nadie pensó antes, pero que estén influyendo en la percepción humana.
- Para evaluar de forma más exhaustiva el rendimiento de CGS, se compararon sus resultados con los resultados de los métodos más habituales en el estado del arte, utilizando en todos los casos características de GoogleNet y ResNet. El mejor resultado de los métodos del estado del arte fue el de SVM con ResNet, con un valor AUROC = 0,82, que no alcanza el valor AUROC = 0,929 de CGS.
- Finalmente, las propuestas anteriores fueron validadas en el marco de una aplicación comercial real. Se demostró un aumento significativo del impacto de los anuncios inmobiliarios online al presentar imágenes con un mayor valor estético. En concreto, se obtuvo una media de incremento de los clics en los anuncios de un 52,54% para los portales inmobiliarios y un 40,08% para las agencias. Estos resultados sugieren que una herramienta que permita ordenar los anuncios en función del valor estético de sus imágenes puede ser de gran utilidad tanto para el usuario como para la empresa o anunciante. Por un lado, mejora la experiencia del usuario al mostrar en las primeras posiciones los anuncios más atractivos desde el punto de vista visual y, por otro lado, consigue atraer y retener en mayor medida al cliente, lo que supone un aumento potencial de la venta para la empresa o anunciante.

8.2. Trabajo futuro

Esta tesis propone un nuevo enfoque para la evaluación estética de imágenes digitales que ha demostrado su eficacia en una aplicación práctica. Sin embargo, surgen también nuevas cuestiones que abordar en desarrollos futuros.

En primer lugar, serán necesarios más estudios para evaluar los resultados obtenidos al aplicar las nuevas CNN de última generación. En cada experimento

de este estudio, solo se ha utilizado una CNN (GoogleNet o ResNet). La capacidad de CGS de filtrar métricas permite probar la combinación de métricas de diferentes fuentes. Así, probaremos una combinación de CNNs y métricas ad-hoc para buscar mejores resultados. La implementación actual no es capaz de procesar un vector de características de entrada con más de 3.000 elementos, por lo que para poder implementar varias fuentes se está refactorizando CGS a un modelo jerárquico multinivel basado en islas que garantice la consecución de mínimos globales y búsquedas con grandes poblaciones.

La estructura distribuida permitirá una división en subpoblaciones con toda la información de entrada posible. El modelo se ejecutará en cada una de ellas en paralelo con intercambios regulares de individuos entre estas subpoblaciones, según una cantidad preestablecida de poblaciones y unos parámetros específicos. El objetivo principal de este enfoque es reinyectar la diversidad periódicamente minimizando las posibilidades de que el modelo final pueda converger hacia un óptimo local.

Por otra parte, en esta tesis doctoral se ha trabajado y estudiado sobre un valor estético que pertenece a la media global de una población. Sin embargo, tal como se explicaba en el estado del arte, la percepción estética es muy subjetiva. Varias personas pueden tener ideas muy diferentes respecto al valor estético de un mismo elemento. Así, se podría decir que cada individuo tiene su propia *huella estética*. Partiendo de esta premisa, se plantea la posibilidad de crear un sistema capaz de predecir el gusto estético de un individuo en particular. Este enfoque podría aportar un mayor beneficio en su aplicación al comercio electrónico y podría mejorar en mayor medida la experiencia del usuario, ya que se le presentarían en primer lugar los productos que más se ajusten a su gusto estético en particular.

Esta línea ya se ha iniciado con la estancia de investigación realizada en el EVALab de la Universidad de Viena, bajo la supervisión de Helmut Leder. Actualmente, se encuentra en fase de experimentación un trabajo en el que intentamos predecir, utilizando CGS, cuán atractivo le parece a un individuo un rostro humano. En trabajos futuros, se pretende seguir estudiando esta hipótesis con imágenes generalistas.

Capítulo 9

Conclusions and future work

This chapter presents the general conclusions obtained in this doctoral thesis and introduces the work that is planned to be developed in relation to this line of research in the near future.

9.1. Conclusions

This thesis addresses different recurrent problems in the field of computational aesthetics and proposes new solutions that are finally validated in a real case study. First, this thesis studies the limitations of the datasets used in computational aesthetics and proposes a new methodology for the creation of image sets that can be applied in Machine Learning problems. Then, a new hybrid genetic algorithm for aesthetics prediction previously developed in the Redes de Neuronas Artificiales y Sistemas Adaptativos (RNASA) group at the University of A Coruña, by Carballal Pazos Perez et al. [136], is presented. Transfer learning was used with this new genetic algorithm and its results were compared with other state-of-the-art models. Finally, both the dataset creation methodology first proposed and the hybrid model that gave the best results in the experimental phase were applied to a real case study. The results suggest that using these tools can improve both user experience and e-commerce performance.

In more detail, the conclusions that can be drawn from this doctoral thesis are the following:

- After detecting some limitations in the datasets used in the state of the art in computational aesthetics, a new methodology for the construction

of datasets was proposed with the aim of improving the training quality of the models. There are two fundamental aspects in the proposal of this methodology compared to the state-of-the-art datasets: first, the images are not binary classified, but distributed by ranks in a score between 0 and 10, and each rank has the same number of images, as opposed to most of the existing datasets that follow a normal distribution; and, second, the images are evaluated a second time by a group of humans in controlled experimental conditions.

- Following the proposed methodology, a new dataset of 1,000 images from the DPChallenge photo portal is also presented, with three different evaluations: (i) the average score obtained in DPChallenge, with at least 100 ratings per image; (ii) the evaluation of the aesthetic value carried out under controlled experimental conditions with a minimum of 10 votes per image; and (iii) the evaluation of the quality of the images carried out under the same conditions as the aesthetic evaluation.
- To test the effectiveness of the above proposals, different state-of-the-art computational methods were used and trained to predict the three criteria (CPDchallenge, aesthetics and quality), resulting in correlations of 0.578, 0.456 and 0.539 respectively, using Spearman's ρ as a measure of performance, and 0.574, 0.451 and 0.562, using Pearson's r . These results suggest that machine learning techniques are better able to learn human perception of technical quality than more subjective aspects such as aesthetics.
- Having addressed the training set issue, a new model for aesthetic measurement based on machine learning, called Correlation by Genetic Search (CGS), was proposed. The results demonstrate that combinations of CNN transfer learning features and the hybrid machine learning method can achieve robust and statistically better results than the state of the art. Using this new method with the GoogleNet network, an AUROC value = 0.929 was achieved for the aesthetic prediction task, which is higher than the AUROC value = 0.896 of the state of the art.
- The results obtained with CGS were compared with the best state-of-the-art results and it was found that the proposed approach outperformed the state-of-the-art in both classification and prediction. One of the reasons for

this may be that by using a set of metrics not designed by humans, it is possible to capture features of the image that no one thought of before, but which are influencing human perception.

- To further evaluate the performance of CGS, its results were compared with the results of the most common methods in the state of the art, using GoogleNet and ResNet features in all cases. The best result of the state-of-the-art methods was that of SVM with ResNet, with an AUROC value = 0.82, which does not reach the AUROC value = 0.929 of CGS.
- Finally, the previous proposals were validated in the framework of a real commercial application. A significant increase in the impact of online real estate ads was demonstrated by presenting images with a higher aesthetic value. Specifically, the average increase in ad clicks was 52.54% for real estate portals and 40.08% for agencies. These results suggest that a tool that allows ads to be sorted according to the aesthetic value of their images can be very useful for both the user and the company or advertiser. On the one hand, it improves the user experience by showing the most visually appealing ads in the top positions and, on the other hand, it attracts and retains the customer to a greater extent, which means a potential increase in sales for the company or advertiser.

9.2. Future work

This thesis proposes a new approach to the aesthetic evaluation of digital images that has proven effective in a practical application. However, there are also new issues to be addressed in future developments.

First, further studies will be necessary to evaluate the results obtained by applying the new state-of-the-art CNNs. In each experiment of this study, only one CNN (GoogleNet or ResNet) has been used. The ability of CGS to filter metrics allows us to test the combination of metrics from different sources. Thus, we will test a combination of CNNs and ad-hoc metrics to look for better results. The current implementation is not able to process an input feature vector with more than 3,000 items, so in order to implement multiple sources, CGS is being refactored to a multilevel hierarchical model based on islands to ensure that global minima and searches with large populations can be achieved.

The distributed structure will allow a division into subpopulations with all the possible input information. The model will run in each of them in parallel with regular exchanges of individuals among these subpopulations, according to a pre-established amount of populations and some specific parameters. The main goal of this approach is to re-inject diversity periodically by minimizing the chances that the final model may converge towards a local optimum.

On the other hand, in this doctoral thesis we have worked and studied an aesthetic value that belongs to the global average of a population. However, as explained in the state of the art, aesthetic perception is very subjective. Different people can have very different ideas about the aesthetic value of the same element. Thus, it could be said that each individual has his or her own aesthetic footprint. Based on this premise, the possibility is raised of creating a system capable of predicting the aesthetic taste of a particular individual. This approach could be of greater benefit in its application to e-commerce and could further enhance the user experience, as the products that most closely match their particular aesthetic taste would be presented to them first.

This line of research has already started with the research stay at the EVALab of the University of Vienna under the supervision of Helmut Leder. A study is currently in the experimental phase in which we are trying to predict, using CGS, how attractive an individual finds a human face. In future work, we intend to further study this hypothesis with generalist images.

Referencias

- [1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016.
- [2] Christophe Ambroise y Geoffrey J McLachlan. Selection bias in gene extraction on the basis of microarray gene-expression data. *Proceedings of the national academy of sciences*, 99(10):6562–6566, 2002.
- [3] Rudolf Arnheim. El pensamiento visual. En *El pensamiento visual*, págs. 343–343. 1971.
- [4] Fred Attneave. Physical determinants of the judged complexity of shapes. *Journal of experimental Psychology*, 53(4):221, 1957.
- [5] Tunç Ozan Aydın, Aljoscha Smolic, y Markus Gross. Automated aesthetic analysis of photographic images. *IEEE transactions on visualization and computer graphics*, 21(1):31–42, 2014.
- [6] Yusuf Cihat Aydın, Parham A Mirzaei, y Sanam Akhavannasab. On the relationship between building energy efficiency, aesthetic features and marketability: Toward a novel policy for energy demand reduction. *Energy Policy*, 128:593–606, 2019.
- [7] N Balakrishnan, IS Triantafyllou, y MV Koutras. Nonparametric control charts based on runs and wilcoxon-type rank-sum statistics. *Journal of Statistical Planning and Inference*, 139(9):3177–3192, 2009.
- [8] John Gilbert Beebe-Center y Carroll Cornelius Pratt. A test of birkhoff’s aesthetic measure. *The Journal of General Psychology*, 17(2):339–353, 1937.

-
- [9] Daniel E Berlyne. Aesthetics and psychobiology. *Journal of Aesthetics and Art Criticism*, 31(4), 1973.
- [10] Subhabrata Bhattacharya, Rahul Sukthankar, y Mubarak Shah. A framework for photo-quality assessment and enhancement based on visual aesthetics. En *Proceedings of the 18th ACM international conference on Multimedia*, págs. 271–280. 2010.
- [11] Simone Bianco, Luigi Celona, Paolo Napoletano, y Raimondo Schettini. Predicting image aesthetics with deep learning. En *International Conference on advanced concepts for intelligent vision systems*, págs. 117–125. Springer, 2016.
- [12] George David Birkhoff. Aesthetic measure. En *Aesthetic Measure*. Harvard University Press, 1933.
- [13] Andrew Bishop, Vic Ciesielski, y Karen Trist. Feature construction using genetic programming for classification of images by aesthetic value. En *International Conference on Evolutionary and Biologically Inspired Music and Art*, págs. 62–73. Springer, 2014.
- [14] Peter H Bloch, Frederic F Brunel, y Todd J Arnold. Individual differences in the centrality of visual product aesthetics: Concept and measurement. *Journal of consumer research*, 29(4):551–565, 2003.
- [15] Avrim L Blum y Pat Langley. Selection of relevant features and examples in machine learning. *Artificial intelligence*, 97(1-2):245–271, 1997.
- [16] Verónica Bolón-Canedo, Noelia Sánchez-Marroño, y Amparo Alonso-Betanzos. A review of feature selection methods on synthetic data. *Knowledge and information systems*, 34(3):483–519, 2013.
- [17] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [18] Gilbert Brighouse. Variability in preferences for simple forms. *Psychological Monographs*, 51(5):68, 1939.
- [19] Adrian Carballal, Luz Castro, Rebeca Perez, y João Correia. Detecting bias on aesthetic image datasets. *International Journal of Creative Interfaces and Computer Graphics (IJCICG)*, 5(2):62–74, 2014.

- [20] Adrian Carballal, Luz Castro, Nereida Rodríguez-Fernández, Iria Santos, Antonino Santos, y Juan Romero. Approach to minimize bias on aesthetic image datasets. En *Interface Support for Creativity, Productivity, and Expression in Computer Graphics*, págs. 203–219. IGI Global, 2019.
- [21] Adrian Carballal, Antonino Santos, Juan Romero, Penousal Machado, João Correia, y Luz Castro. Distinguishing paintings from photographs by complexity estimates. *Neural Computing and Applications*, 30(6):1957–1969, 2018.
- [22] Luz Castro, Rebeca Perez, Antonino Santos, y Adrian Carballal. Authorship and aesthetics experiments: comparison of results between human and computational systems. En *International Conference on Evolutionary and Biologically Inspired Music and Art*, págs. 74–84. Springer, 2014.
- [23] M Luz Castro. Diseño, construcción e validación dun clasificador de imaxes segundo criterios estéticos. 2017.
- [24] Camilo J Cela-Conde, Francisco J Ayala, Enric Munar, Fernando Maestú, Marcos Nadal, Miguel A Capó, David del Río, Juan J López-Ibor, Tomás Ortiz, Claudio Mirasso, et al. Sex-related similarities and differences in the neural correlates of beauty. *Proceedings of the National Academy of Sciences*, 106(10):3847–3852, 2009.
- [25] Camilo J Cela-Conde, Gisèle Marty, Fernando Maestú, Tomás Ortiz, Enric Munar, Alberto Fernández, Miquel Roca, Jaume Rosselló, y Felipe Quesney. Activation of the prefrontal cortex in the human visual aesthetic perception. *Proceedings of the National Academy of Sciences*, 101(16):6321–6325, 2004.
- [26] Andrew T Chamberlain. On the evolution of human aesthetic preferences. *Assemblage: University of Sheffield Graduate Student Journal of Archaeology*, 5, 2000.
- [27] Ming Chen y Jan Allebach. Aesthetic quality inference for online fashion shopping. En *Imaging and Multimedia Analytics in a Web and Mobile World 2014*, tomo 9027, págs. 6–12. SPIE, 2014.
- [28] Susan F Chipman. Complexity and structure in visual patterns. *Journal of Experimental Psychology: General*, 106(3):269, 1977.

- [29] Susan F Chipman y Morton J Mendelson. Influence of six types of visual structure on complexity judgments in children and adults. *Journal of Experimental Psychology: Human Perception and Performance*, 5(2):365, 1979.
- [30] François Chollet et al. keras. 2015.
- [31] Stergios Christodoulidis, Marios Anthimopoulos, Lukas Ebner, Andreas Christe, y Stavroula Mougiakakou. Multisource transfer learning with convolutional neural networks for lung pattern analysis. *IEEE journal of biomedical and health informatics*, 21(1):76–84, 2016.
- [32] Vic Ciesielski, Perry Barile, y Karen Trist. Finding image features associated with high aesthetic value by machine learning. En *International Conference on Evolutionary and Biologically Inspired Music and Art*, págs. 47–58. Springer, 2013.
- [33] Juan Cordero Ruiz. Sobre la percepción visual y leyes de la gestalt. https://issuu.com/luly/docs/juan_cordero_ruiz_leyes_gestalt, 2013. Consulta: 03/06/2021.
- [34] Gabriella Csurka, Christopher Dance, Lixin Fan, Jutta Willamowski, y Cédric Bray. Visual categorization with bags of keypoints. En *Workshop on statistical learning in computer vision, ECCV*, tomo 1, págs. 1–2. Prague, 2004.
- [35] Ritendra Datta, Dhiraj Joshi, Jia Li, y James Z Wang. Studying aesthetics in photographic images using a computational approach. En *European conference on computer vision*, págs. 288–301. Springer, 2006.
- [36] Ritendra Datta, Jia Li, y James Z Wang. Algorithmic inferencing of aesthetics and emotion in natural images: An exposition. En *2008 15th IEEE international conference on image processing*, págs. 105–108. IEEE, 2008.
- [37] Ritendra Datta y James Z Wang. Acquine: aesthetic quality inference engine-real-time automatic rating of photo aesthetics. En *Proceedings of the international conference on Multimedia information retrieval*, págs. 421–424. 2010.
- [38] Roland Clark Davis. An evaluation and test of birkhoff’s aesthetic measure formula. *The Journal of General Psychology*, 15(2):231–240, 1936.

- [39] Soma Debnath y Suvamoy Changder. Computational approaches to aesthetic quality assessment of digital photographs: state of the art and future research directives. *Pattern Recognition and Image Analysis*, 30(4):593–606, 2020.
- [40] Liqiong Deng y Marshall Scott Poole. Aesthetic design of e-commerce web pages—webpage complexity, order and preference. *Electronic Commerce Research and Applications*, 11(4):420–440, 2012.
- [41] Yubin Deng, Chen Change Loy, y Xiaoou Tang. Image aesthetic assessment: An experimental survey. *IEEE Signal Processing Magazine*, 34(4):80–106, 2017.
- [42] Pierre A Devijver y Josef Kittler. *Pattern recognition: A statistical approach*. Prentice hall, 1982.
- [43] David M Diez, Christopher D Barr, y Mine Cetinkaya-Rundel. *OpenIntro statistics*. OpenIntro Boston, MA, USA:, 2012.
- [44] Hans Jurgen Eysenck. ‘type’-factors in aesthetic judgements. *British Journal of Psychology. General Section*, 31(3):262–270, 1941.
- [45] Hans Jurgen Eysenck. The experimental study of the ‘good gestalt’—a new approach. *Psychological Review*, 49(4):344, 1942.
- [46] HJ Eysenck y M Castle. Comparative study of artists and nonartists on the maitland graves design judgment test. *Journal of Applied Psychology*, 55(4):389, 1971.
- [47] Gustav Theodor Fechner. *Vorschule der aesthetik*, tomo 1. Druck und Verlag von Breitkopf & Härtel, 1876.
- [48] Carlos Fernandez-Lozano, Jose A Seoane, Marcos Gestal, Tom R Gaunt, Julian Dorado, y Colin Campbell. Texture classification using feature selection and kernel-based techniques. *Soft Computing*, 19(9):2469–2480, 2015.
- [49] Alex Forsythe, Marcos Nadal, Noel Sheehy, Camilo J Cela-Conde, y Martin Sawey. Predicting beauty: fractal dimension and visual complexity in art. *British journal of psychology*, 102(1):49–70, 2011.

-
- [50] Robert Francés. *Psicología del arte y la estética*, tomo 3. Ediciones AKAL, 2005.
- [51] Eibe Frank, Yong Wang, Stuart Inglis, Geoffrey Holmes, y Ian H Witten. Using model trees for classification. *Machine learning*, 32(1):63–76, 1998.
- [52] Cynthia Freeland. *Pero ¿esto es arte?: una introducción a la teoría del arte*. Ediciones Cátedra, 2001.
- [53] Jerome Friedman, Trevor Hastie, y Robert Tibshirani. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The annals of statistics*, 28(2):337–407, 2000.
- [54] Jerome H Friedman. Stochastic gradient boosting. *Computational statistics & data analysis*, 38(4):367–378, 2002.
- [55] Charlie D Frowd, Peter JB Hancock, y Derek Carson. Evofit: A holistic, evolutionary facial imaging technique for creating composites. *ACM Transactions on applied perception (TAP)*, 1(1):19–39, 2004.
- [56] Salvador García, Alberto Fernández, Julián Luengo, y Francisco Herrera. Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power. *Information sciences*, 180(10):2044–2064, 2010.
- [57] Stephen R Garner et al. Weka: The waikato environment for knowledge analysis. En *Proceedings of the New Zealand computer science research students conference*, tomo 1995, págs. 57–64. 1995.
- [58] Peter Gehler y Sebastian Nowozin. On feature combination for multiclass object classification. En *2009 IEEE 12th International Conference on Computer Vision*, págs. 221–228. IEEE, 2009.
- [59] Aurélien Géron. Hands-on machine learning with scikit-learn and tensorflow: Concepts. *Tools, and Techniques to build intelligent systems*, 2017.
- [60] Mohsen Ghafoorian, Alireza Mehrtash, Tina Kapur, Nico Karssemeijer, Elena Marchiori, Mehran Pesteie, Charles RG Guttman, Frank-Erik de Leeuw, Clare M Tempany, Bram van Ginneken, et al. Transfer learning for

- domain adaptation in mri: Application in brain lesion segmentation. En *International conference on medical image computing and computer-assisted intervention*, págs. 516–524. Springer, 2017.
- [61] Maitland Graves. Design judgment test. 1948.
- [62] Maitland E. Graves. *The art of color and design*. McGraw-Hill, 1951.
- [63] Brandon Greenwell, Bradley Boehmke, Jay Cunningham, y Developers GBM. Generalized boosted regression models. <https://CRAN.R-project.org/package=gbm>, 2020. Consulta: 12/12/2021.
- [64] Derya GÜLÇİN. Predicting visual aesthetic preferences of landscapes near historical sites by fluency theory using social media data and gis. *International Journal of Geography and Geography Education*, (43):265–277, 2021.
- [65] Guanjun Guo, Hanzi Wang, Chunhua Shen, Yan Yan, y Hong-Yuan Mark Liao. Automatic image cropping for visual aesthetic enhancement using deep neural networks and cascaded regression. *IEEE Transactions on Multimedia*, 20(8):2073–2085, 2018.
- [66] Isabelle Guyon, Steve Gunn, Masoud Nikraves, y Lofti A Zadeh. *Feature extraction: foundations and applications*, tomo 207. Springer, 2008.
- [67] Isabelle Guyon, Jason Weston, Stephen Barnhill, y Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1):389–422, 2002.
- [68] AC Hall. Measures of the complexity of random black and white and coloured stimuli. *Perceptual and motor skills*, 29(3):773–774, 1969.
- [69] Kaiming He, Xiangyu Zhang, Shaoqing Ren, y Jian Sun. Deep residual learning for image recognition. En *Proceedings of the IEEE conference on computer vision and pattern recognition*, págs. 770–778. 2016.
- [70] Klaus Hechenbichler y Klaus Schliep. Weighted k-nearest-neighbor techniques and ordinal classification. *Collaborative Research Center 386*, 2004.

- [71] Yong-Lian Hui, John See, Magzhan Kairanbay, y Lai-Kuan Wong. Multigap: Multi-pooled inception network with text augmentation for aesthetic prediction of photographs. En *2017 IEEE International Conference on Image Processing (ICIP)*, págs. 1722–1726. IEEE, 2017.
- [72] Ronald R Hocking. A biometrics invited paper. the analysis and selection of variables in linear regression. *Biometrics*, págs. 1–49, 1976.
- [73] Menno Hogeboom y Cees van Leeuwen. Visual search strategy and perceptual organization covary with individual preference and structural complexity. *Acta Psychologica*, 95(2):141–164, 1997.
- [74] Mark J Huiskes y Michael S Lew. The mir flickr retrieval evaluation. En *Proceedings of the 1st ACM international conference on Multimedia information retrieval*, págs. 39–43. 2008.
- [75] Mark J Huiskes, Bart Thomee, y Michael S Lew. New trends and ideas in visual concept detection: The mir flickr retrieval evaluation initiative. En *Proceedings of the international conference on Multimedia information retrieval*, págs. 527–536. 2010.
- [76] Clifford M Hurvich y Chih—Ling Tsai. The impact of model selection on inference in linear regression. *The American Statistician*, 44(3):214–217, 1990.
- [77] Kiyohito Iigaya, Sanghyun Yi, Iman A Wahle, Koranis Tanwisuth, y John P O’Doherty. Aesthetic preference for art can be predicted from a mixture of low-and high-level visual features. *Nature human behaviour*, 5(6):743–755, 2021.
- [78] Azlan Iqbal. Computational aesthetics. <https://www.britannica.com/topic/computational-aesthetics>. Consulta: 20/05/2022.
- [79] Xin Jin, Le Wu, Xiaodong Li, Xiaokun Zhang, Jingying Chi, Siwei Peng, Shiming Ge, Geng Zhao, y Shuying Li. Ilnet: inception modules with connected local and global features for efficient image aesthetic quality classification using domain adaptation. *IET Computer Vision*, 13(2):206–212, 2019.

-
- [80] Xin Jin, Le Wu, Geng Zhao, Xinghui Zhou, Xiaokun Zhang, y Xiaodong Li. Idea: A new dataset for image aesthetic scoring. *Multimedia Tools and Applications*, 79(21):14341–14355, 2020.
- [81] Victor S Johnston. Facial beauty and mate choice decisions. *Trends in Cognitive Science*, 10(1):9–13, 2005.
- [82] Adrian Kaehler y Gary Bradski. *Learning OpenCV 3: computer vision in C++ with the OpenCV library*. “O’Reilly Media, Inc.”, 2016.
- [83] Saira Kanwal, Muhammad Uzair, y Habib Ullah. A survey of hand crafted and deep learning methods for image aesthetic assessment. *arXiv preprint arXiv:2103.11616*, 2021.
- [84] Yueying Kao, Ran He, y Kaiqi Huang. Deep aesthetic quality assessment with semantic information. *IEEE Transactions on Image Processing*, 26(3):1482–1495, 2017.
- [85] Yueying Kao, Kaiqi Huang, y Steve Maybank. Hierarchical aesthetic quality assessment using deep convolutional neural networks. *Signal Processing: Image Communication*, 47:500–510, 2016.
- [86] SMR Kazemi, Behrouz Minaei Bidgoli, Shahaboddin Shamshirband, Seyyed Mehdi Karimi, Mohammad Ali Ghorbani, Kwok-wing Chau, y Reza Kazem Pour. Novel genetic-based negative correlation learning for estimating soil temperature. *Engineering Applications of Computational Fluid Mechanics*, 12(1):506–516, 2018.
- [87] Yan Ke, Xiaoou Tang, y Feng Jing. The design of high-level features for photo quality assessment. En *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06)*, tomo 1, págs. 419–426. IEEE, 2006.
- [88] Robert H Knapp y Samuel Green. Preferences for styles of abstract art and their personality correlates. *Journal of Projective Techniques*, 24(4):396–402, 1960.
- [89] Robert H Knapp, Lowell R McElroy, y James Vaughn. On blithe and melancholic aestheticism. *The Journal of General Psychology*, 67(1):3–10, 1962.

- [90] Shu Kong, Xiaohui Shen, Zhe Lin, Radomir Mech, y Charless Fowlkes. Photo aesthetics ranking network with attributes and content adaptation. En *European conference on computer vision*, págs. 662–679. Springer, 2016.
- [91] Carolyn Korsmeyer. *Aesthetics: The Big Questions*. Wiley-Blackwell, 1998.
- [92] Alex Krizhevsky, Ilya Sutskever, y Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.
- [93] William H Kruskal y W Allen Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260):583–621, 1952.
- [94] Mineichi Kudo y Jack Sklansky. A comparative evaluation of medium-and large-scale feature selectors for pattern classifiers. *Kybernetika*, 34(4):429–434, 1998.
- [95] Hui-Jin Lee, Ki-Sang Hong, Henry Kang, y Seungyong Lee. Photo aesthetics analysis via dcnn feature encoding. *IEEE Transactions on Multimedia*, 19(8):1921–1932, 2017.
- [96] Congcong Li y Tsuhan Chen. Aesthetic visual quality assessment of paintings. *IEEE Journal of selected topics in Signal Processing*, 3(2):236–252, 2009.
- [97] Yuxin Li, Yuanyuan Pu, Dan Xu, Wenhua Qian, y Lipeng Wang. Image aesthetic quality evaluation using convolution neural network embedded fine-tune. En *CCF Chinese Conference on Computer Vision*, págs. 269–283. Springer, 2017.
- [98] Rensis Likert. A technique for the measurement of attitudes. *Archives of psychology*, 1932.
- [99] Huan Liu y Hiroshi Motoda. Feature transformation and subset selection. *IEEE Intell Syst Their Appl*, 13(2):26–28, 1998.
- [100] Ligang Liu, Renjie Chen, Lior Wolf, y Daniel Cohen-Or. Optimizing photo composition. En *Computer graphics forum*, tomo 29, págs. 469–478. Wiley Online Library, 2010.

- [101] Wei Liu y Sanjay Chawla. Class confidence weighted knn algorithms for imbalanced data sets. En *Pacific-Asia conference on knowledge discovery and data mining*, págs. 345–356. Springer, 2011.
- [102] Xin Lu, Zhe Lin, Xiaohui Shen, Radomir Mech, y James Z Wang. Deep multi-patch aggregation network for image style, aesthetics, and quality estimation. En *Proceedings of the IEEE international conference on computer vision*, págs. 990–998. 2015.
- [103] Wei Luo, Xiaogang Wang, y Xiaoou Tang. Content-based photo quality assessment. En *2011 international conference on computer vision*, págs. 2206–2213. IEEE, 2011.
- [104] Yiwen Luo y Xiaoou Tang. Photo and video quality evaluation: Focusing on the subject. En *European Conference on Computer Vision*, págs. 386–399. Springer, 2008.
- [105] Fernando Jorge Penousal Martins Machado. *Inteligencia artificial e arte*. Tesis Doctoral, 2007.
- [106] Penousal Machado y Amílcar Cardoso. Computing aesthetics. En *Brazilian symposium on artificial intelligence*, págs. 219–228. Springer, 1998.
- [107] Penousal Machado, Juan Romero, Marcos Nadal, Antonino Santos, João Correia, y Adrián Carballal. Computerized measures of visual complexity. *Acta Psychologica*, 160:43–57, 2015. ISSN 0001-6918. doi:<https://doi.org/10.1016/j.actpsy.2015.06.005>. URL <https://www.sciencedirect.com/science/article/pii/S0001691815300160>.
- [108] Penousal Machado, Juan Romero, Antonino Santos, Amílcar Cardoso, y Alejandro Pazos. On the development of evolutionary artificial artists. *Computers & Graphics*, 31(6):818–826, 2007.
- [109] Long Mai, Hailin Jin, y Feng Liu. Composition-preserving deep photo aesthetics assessment. En *Proceedings of the IEEE conference on computer vision and pattern recognition*, págs. 497–506. 2016.
- [110] Sebastián Maldonado, Richard Weber, y Jayanta Basak. Simultaneous feature selection and classification using kernel-penalized support vector machines. *Information Sciences*, 181(1):115–128, 2011.

- [111] Jean-Claude Malela-Majika. New distribution-free memory-type control charts based on the wilcoxon rank-sum statistic. *Quality Technology & Quantitative Management*, 18(2):135–155, 2021.
- [112] Simon Malesev y Michael Cherry. Digital and social media marketing-growing market share for construction smes. *Construction Economics and Building*, 21(1):65–82, 2021.
- [113] Pratap Chandra Mandal. Changing nature of customer relationships: strategies and initiatives to beat the competition. *International Journal of Competitiveness*, 2(1):1–16, 2021.
- [114] Luca Marchesotti, Florent Perronnin, Diane Larlus, y Gabriela Csurka. Assessing the aesthetic quality of photographs using generic image descriptors. En *2011 international conference on computer vision*, págs. 1784–1791. IEEE, 2011.
- [115] Manuela M Marin y Helmut Leder. Examining complexity across domains: relating subjective and objective measures of affective environmental scenes, paintings and music. *PloS one*, 8(8):e72412, 2013.
- [116] José Luis Martorell y José Luis Prieto. Manual fundamentos de la psicología. *Centro de Estudios Ramón Areces. Colección de Psicología. España*, 2002.
- [117] Keith E Mathias, L Darrell Whitley, Christof Stork, y Tony Kusuma. Staged hybrid genetic search for seismic data imaging. En *Proceedings of the First IEEE Conference on Evolutionary Computation. IEEE World Congress on Computational Intelligence*, págs. 356–361. IEEE, 1994.
- [118] Daniel F McCaffrey, Beth Ann Griffin, Daniel Almirall, Mary Ellen Slaughter, Rajeev Ramchand, y Lane F Burgette. A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Statistics in medicine*, 32(19):3388–3414, 2013.
- [119] Jon McCormack y Andy Lomas. Deep learning of individual aesthetics. *Neural Computing and Applications*, 33(1):3–17, 2021.
- [120] John H McDonald. *Handbook of biological statistics*, tomo 2. sparky house publishing Baltimore, MD, 2009.

- [121] Norman Charles Meier. Art in human affairs; an introduction to the psychology of art. 1942.
- [122] Afonso Menegola, Michel Fornaciali, Ramon Pires, Flávia Vasques Bittencourt, Sandra Avila, y Eduardo Valle. Knowledge transfer for melanoma screening with deep learning. En *2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017)*, págs. 297–300. IEEE, 2017.
- [123] Anders Pape Møller. Developmental stability and fitness: a review. *The American Naturalist*, 149(5):916–932, 1997.
- [124] Andreas C Müller y Sarah Guido. *Introduction to machine learning with Python: a guide for data scientists*. .O'Reilly Media, Inc.", 2016.
- [125] Caitlin Mullin, Gregor Hayn-Leichsenring, Christoph Redies, y Johan Wagemans. The gist of beauty: an investigation of aesthetic perception in rapidly presented images. *Electronic Imaging*, 2017(14):248–256, 2017.
- [126] Thomas Munro. Evolution in the arts, and other theories of culture history. *Philosophy*, 40(153), 1965.
- [127] Naila Murray, Luca Marchesotti, y Florent Perronnin. Ava: A large-scale database for aesthetic visual analysis. En *2012 IEEE conference on computer vision and pattern recognition*, págs. 2408–2415. IEEE, 2012.
- [128] Marcos Nadal, Enric Munar, Gisèle Marty, y Camilo José Cela-Conde. Visual complexity and beauty appreciation: Explaining the divergence of results. *Empirical Studies of the Arts*, 28(2):173–191, 2010.
- [129] Marcos Nadal Roberts. Complexity and aesthetic preference for diverse visual stimuli. *Doctoral*), *Universitat de les Illes Balears, Palma, Spain*, 2007.
- [130] Preeti Nayal, Neeraj Pandey, y Justin Paul. Covid-19 pandemic and consumer-employee-organization wellbeing: A dynamic capability theory approach. *Journal of Consumer Affairs*, 56(1):359–390, 2022.
- [131] L Neumann, M Sbert, B Gooch, W Purgathofer, et al. Defining computational aesthetics. *Computational aesthetics in graphics, visualization and imaging*, págs. 13–18, 2005.

- [132] Jerzy Neyman. Basic ideas and some recent results of the theory of testing statistical hypotheses. *Journal of the Royal statistical society*, 105(4):292–327, 1942.
- [133] Masashi Nishiyama, Takahiro Okabe, Imari Sato, y Yoichi Sato. Aesthetic quality classification of photographs based on color harmony. En *CVPR 2011*, págs. 33–40. IEEE, 2011.
- [134] Nikita Orlov, Lior Shamir, Tomasz Macura, Josiah Johnston, D Mark Eckley, y Ilya G Goldberg. Wnd-charm: Multi-purpose image classification using compound image transforms. *Pattern recognition letters*, 29(11):1684–1693, 2008.
- [135] Sinno Jialin Pan y Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.
- [136] Rafael Ivan Pazos Perez, Adrian Carballal, Juan R Rabuñal, Omar A Mures, y María D García-Vidaurrázaga. Predicting vertical urban growth using genetic evolutionary algorithms in tokyo’s minato ward. *Journal of Urban Planning and Development*, 144(1):04017024, 2017.
- [137] Karl Pearson. Notes on the history of correlation. *Biometrika*, 13(1):25–45, 1920.
- [138] Cameron Pugach, Helmut Leder, y Daniel J Graham. How stable are human aesthetic preferences across the lifespan? *Frontiers in human neuroscience*, 11:289, 2017.
- [139] Juan Romero, Penousal Machado, Adrian Carballal, y Olga Osorio. Aesthetic classification and sorting based on image compression. En *European Conference on the Applications of Evolutionary Computation*, págs. 394–403. Springer, 2011.
- [140] Juan Romero, Penousal Machado, Adrian Carballal, y Antonino Santos. Using complexity estimates in aesthetic image classification. *Journal of Mathematics and the Arts*, 6(2-3):125–136, 2012.
- [141] EE Rump. Is there a general factor of preference for complexity? *Perception & Psychophysics*, 3(5):346–348, 1968.

- [142] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.
- [143] Yvan Saeys, Inaki Inza, y Pedro Larranaga. A review of feature selection techniques in bioinformatics. *bioinformatics*, 23(19):2507–2517, 2007.
- [144] Curtis A Samuels, George Butterworth, Tony Roberts, Lida Graupner, y Graham Hole. Facial aesthetics: Babies prefer attractiveness to symmetry. *Perception*, 23(7):823–831, 1994.
- [145] Iria Santos, Luz Castro, Nereida Rodriguez-Fernandez, Alvaro Torrente-Patino, y Adrian Carballal. Artificial neural networks and deep learning in the visual arts: A review. *Neural Computing and Applications*, 33(1):121–157, 2021.
- [146] Pakizar Shamoii, Atsushi Inoue, y Hiroharu Kawanaka. Modeling aesthetic preferences: Color coordination and fuzzy sets. *Fuzzy Sets and Systems*, 395:217–234, 2020.
- [147] Ali Sharif Razavian, Hossein Azizpour, Josephine Sullivan, y Stefan Carlsson. Cnn features off-the-shelf: an astounding baseline for recognition. En *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, págs. 806–813. 2014.
- [148] Kekai Sheng, Weiming Dong, Haibin Huang, Menglei Chai, Yong Zhang, Chongyang Ma, y Bao-Gang Hu. Learning to assess visual aesthetics of food images. *Computational Visual Media*, 7(1):139–152, 2021.
- [149] Lars Strother y Michael Kubovy. Perceived complexity and the grouping effect in band patterns. *Acta Psychologica*, 114(3):229–244, 2003.
- [150] Hsiao-Hang Su, Tse-Wei Chen, Chieh-Chi Kao, Winston H Hsu, y Shao-Yi Chien. Scenic photo quality assessment with bag of aesthetics-preserving features. En *Proceedings of the 19th ACM international conference on Multimedia*, págs. 1213–1216. 2011.

- [151] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, y Andrew Rabinovich. Going deeper with convolutions. En *Proceedings of the IEEE conference on computer vision and pattern recognition*, págs. 1–9. 2015.
- [152] Hippolyte Taine. *Filosofía del arte*. F. Sempere y Compañía, 1905.
- [153] Yunlan Tan, Yimin Zhou, Guangyao Li, y Anmin Huang. Computational aesthetics of photos quality assessment based on improved artificial neural network combined with an autoencoder technique. *Neurocomputing*, 188:50–62, 2016.
- [154] Xiaoou Tang, Wei Luo, y Xiaogang Wang. Content-based photo quality assessment. *IEEE Transactions on Multimedia*, 15(8):1930–1943, 2013.
- [155] Riccardo Taormina, Kwok-Wing Chau, y Bellie Sivakumar. Neural network river forecasting through baseflow separation and binary-coded swarm optimization. *Journal of Hydrology*, 529:1788–1797, 2015.
- [156] Władysław Tatarkiewicz. *Czworakie rozumienie klasyczności*. 1996.
- [157] John Taylor. *Introduction to error analysis, the study of uncertainties in physical measurements*. 1997.
- [158] Katja Thömmes y Ronald Hübner. Why people press “like”: A new measure for aesthetic appeal derived from instagram data. *Psychology of Aesthetics, Creativity, and the Arts*, 2020.
- [159] Xinmei Tian, Zhe Dong, Kuiyuan Yang, y Tao Mei. Query-dependent aesthetic model with deep learning for photo quality assessment. *IEEE Transactions on Multimedia*, 17(11):2035–2048, 2015.
- [160] Nguyen Hoang Tien, Pham Thi Diem, Nguyen Thanh Vu, Ho Tien Dung, Bui Xuan Bien, Phan Minh Duc, y Nguyen Van Dat. Customer care and customer relationship maintenance at ministop, familymart and coopsmile in vietnam. *International Journal of Research and Growth Evaluation*, 2(4), 2021.
- [161] Hanghang Tong, Mingjing Li, Hong-Jiang Zhang, Jingrui He, y Changshui Zhang. Classification of digital photos taken by photographers or home

- users. En *Pacific-Rim Conference on Multimedia*, págs. 198–205. Springer, 2004.
- [162] Sumit Tripathi y Neeraj Sharma. Computer-aided automatic approach for denoising of magnetic resonance images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 9(6):707–716, 2021.
- [163] Georgia Tsiliki, Cristian R Munteanu, Jose A Seoane, Carlos Fernandez-Lozano, Haralambos Sarimveis, y Egon L Willighagen. Rregrs: an r package for computer-aided model selection with multiple regression models. *Journal of cheminformatics*, 7(1):1–16, 2015.
- [164] Jianyu Wang y Jan Allebach. Automatic assessment of online fashion shopping photo aesthetic quality. En *2015 IEEE International Conference on Image Processing (ICIP)*, págs. 2915–2919. IEEE, 2015.
- [165] Shui-Hua Wang, Junding Sun, Preetha Phillips, Guihu Zhao, y Yu-Dong Zhang. Polarimetric synthetic aperture radar image segmentation by convolutional neural network using graphical processing units. *Journal of Real-Time Image Processing*, 15(3):631–642, 2018.
- [166] Shui-Hua Wang, Chaosheng Tang, Junding Sun, Jingyuan Yang, Chenxi Huang, Preetha Phillips, y Yu-Dong Zhang. Multiple sclerosis identification by 14-layer convolutional neural network with batch normalization, dropout, and stochastic pooling. *Frontiers in neuroscience*, 12:818, 2018.
- [167] Weining Wang, Dong Cai, Li Wang, Qinghua Huang, Xiangmin Xu, y Xuelong Li. Synthesized computational aesthetic evaluation of photos. *Neurocomputing*, 172:244–252, 2016.
- [168] CO Weber. The aesthetics of rectangles and theories of affection. *Journal of Applied Psychology*, 15(3):310, 1931.
- [169] Frank Wilcoxon. Individual comparisons by ranking methods. En *Breakthroughs in statistics*, págs. 196–202. Springer, 1992.
- [170] Lai-Kuan Wong y Kok-Lim Low. Saliency-enhanced image aesthetics class prediction. En *2009 16th IEEE international conference on image processing (ICIP)*, págs. 997–1000. IEEE, 2009.

-
- [171] Xiaoran Wu. Interpretable aesthetic analysis model for intelligent photography guidance systems. En *27th International Conference on Intelligent User Interfaces*, págs. 661–671. 2022.
- [172] Yingjie Xia, Zhenguang Liu, Yan Yan, Yanxiang Chen, Luming Zhang, y Roger Zimmermann. Media quality assessment by perceptual gaze-shift patterns discovery. *IEEE Transactions on Multimedia*, 19(8):1811–1820, 2017.
- [173] Mei-Chen Yeh y Chun-Hui Chuang. Assessing the aesthetic quality of photographs through group comparison. En *2014 IEEE International Conference on Consumer Electronics-Taiwan*, págs. 91–92. IEEE, 2014.
- [174] Wenhui Yu, Huidi Zhang, Xiangnan He, Xu Chen, Li Xiong, y Zheng Qin. Aesthetic-based clothing recommendation. En *Proceedings of the 2018 world wide web conference*, págs. 649–658. 2018.
- [175] Hao Zhu, Man-Di Luo, Rui Wang, Ai-Hua Zheng, y Ran He. Deep audiovisual learning: A survey. *International Journal of Automation and Computing*, 18(3):351–376, 2021.
- [176] Xianghui Zhu y Teng-Sheng Moh. Assessing photograph aesthetic quality with color based descriptor. En *2016 International Conference on High Performance Computing & Simulation (HPCS)*, págs. 222–229. IEEE, 2016.

Parte IV
Apéndices

Apéndice A

Producción científica

En este primer apéndice se recoge la producción científica generada durante la etapa predoctoral. Se diferencia entre artículos científicos publicados en revistas de alto impacto y las participaciones en congresos nacionales e internacionales que surgieron en el desarrollo de esta tesis doctoral.

Artículos publicados en revistas JCR

Research Article

Avoiding the Inherent Limitations in Datasets Used for Measuring Aesthetics When Using a Machine Learning Approach

Adrian Carballal ¹, **Carlos Fernandez-Lozano** ^{1,2}, **Nereida Rodriguez-Fernandez**,³
Luz Castro,³ and **Antonino Santos**¹

¹Computer Science Department, Faculty of Computer Science, University of A Coruña, A Coruña 15071, Spain

²Investigación Biomédica de A Coruña (INIBIC), Complejo Hospitalario Universitario de A Coruña (CHUAC), A Coruña 15006, Spain

³Computer Science Department, Faculty of Communication Science, University of A Coruña, A Coruña 15071, Spain

Correspondence should be addressed to Adrian Carballal; adrian.carballal@udc.es

Received 17 August 2018; Revised 14 November 2018; Accepted 20 December 2018; Published 8 January 2019

Guest Editor: Jon McCormack

Copyright © 2019 Adrian Carballal et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

An important topic in evolutionary art is the development of systems that can mimic the aesthetics decisions made by human beings, e.g., fitness evaluations made by humans using interactive evolution in generative art. This paper focuses on the analysis of several datasets used for aesthetic prediction based on ratings from photography websites and psychological experiments. Since these datasets present problems, we proposed a new dataset that is a subset of DPChallenge.com. Subsequently, three different evaluation methods were considered, one derived from the ratings available at DPChallenge.com and two obtained under experimental conditions related to the aesthetics and quality of images. We observed different criteria in the DPChallenge.com ratings, which had more to do with the photographic quality than with the aesthetic value. Finally, we explored learning systems other than state-of-the-art ones, in order to predict these three values. The obtained results were similar to those using state-of-the-art procedures.

1. Introduction

Estimating aesthetic value and the complexity of an image is a technological challenge that has recently been addressed by numerous fields, including psychology and artificial intelligence. Several research groups have attempted to create computer systems that are able to learn the aesthetics perception of a group of human beings as a part of a generative system (such as evolutionary art systems) or that can be used for automatic image selection or ordering. Given the subjective nature of the aesthetic problem, the selection of the dataset for the training is vital. This paper explores a new way to build a dataset and provide initial results by using machine learning techniques.

Previous research studies [1, 2] have concluded that the degree of generalisation of some existing sample sets was not enough to take them as reference in the training of automated prediction and classification of images. Other functional

limitations were identified in these datasets, which are also mentioned in this paper.






In order to solve the problems identified in these datasets, this paper describes the creation of a new set of images from the website DPChallenge.com, with greater statistical consistency. Besides, this new dataset was evaluated in terms of aesthetics and quality by a group of individuals under controlled experimental conditions. This makes it the first dataset evaluated by two different populations (the one evaluating at the DPChallenge.com portal and the one evaluating it in person).

With the new dataset created, several Machine Learning-based models were trained for the automated prediction of the aesthetic and quality value and that of DPChallenge.com.

This paper starts with a state-of-the-art section on the datasets created for the automated prediction and classification of images. In Section 3, the limitations found in such sample sets are provided. Section 4 describes the method for

Article

Comparison of Outlier-Tolerant Models for Measuring Visual Complexity

Adrian Carballal ^{1,2,*} , Carlos Fernandez-Lozano ^{1,2} , Nereida Rodriguez-Fernandez ^{1,3} ,
Iria Santos ^{1,3}  and Juan Romero ^{1,3} 

¹ CITIC-Research Center of Information and Communication Technologies, University of A Coruña, 15071 A Coruña, Spain; carlos.fernandez@udc.es (C.F.-L.); nereida.rodriguez@udc.es (N.R.-F.); Iria.santos@udc.es (I.S.); juan.romero1@udc.es (J.R.)

² Department of Computer Science and Information Technologies, Faculty of Computer Science, University of A Coruña, Campus Elviña s/n, 15071 A Coruña, Spain

³ Department of Computer Science and Information Technologies, Faculty of Communication Science, University of A Coruña, Campus Elviña s/n, 15071 A Coruña, Spain

* Correspondence: adrian.carballal@udc.es

Received: 20 February 2020; Accepted: 23 April 2020; Published: 24 April 2020



Abstract: Providing the visual complexity of an image in terms of impact or aesthetic preference can be of great applicability in areas such as psychology or marketing. To this end, certain areas such as Computer Vision have focused on identifying features and computational models that allow for satisfactory results. This paper studies the application of recent ML models using input images evaluated by humans and characterized by features related to visual complexity. According to the experiments carried out, it was confirmed that one of these methods, Correlation by Genetic Search (CGS), based on the search for minimum sets of features that maximize the correlation of the model with respect to the input data, predicted human ratings of image visual complexity better than any other model referenced to date in terms of correlation, RMSE or minimum number of features required by the model. In addition, the variability of these terms were studied eliminating images considered as outliers in previous studies, observing the robustness of the method when selecting the most important variables to make the prediction.

Keywords: machine learning; visual complexity; visual stimuli; correlation; human-computer interaction; compression error; psychiatry and psychology

1. Introduction

Quantifying the visual complexity of an image is a task that is still very complex for the research community in Computer Vision. In a society full of visual stimuli, having the complexity value of an image could be a valuable contribution to work on visual impact, for example. Research groups from different areas of knowledge have focused on identifying features that are highly correlated with human aesthetic and visual preferences. Research has shown that different perceptual features, such as color, color combination, contour, or symmetry, influence people's visual preferences and affective responses [1–3]. Complexity is considered to have a strong impact on preference and affection, given its relationship with excitement [4,5] and has therefore gained great relevance in psychological models of aesthetic appreciation [4,6]. Certain features of the human mind directly influence the sensation of perceived complexity [7]. Many factors, such as previous exposure to certain visual stimuli or familiarity with the given stimuli, affect the human perception of visual complexity. Authors such as Frijda (1989) or Blood & Zatorre (2001) also state that the aesthetic experience is affective or even emotional [8,9]. Therefore, there is no doubt that the perception of complexity is inevitably



Artificial Neural Networks and Deep Learning in the Visual Arts: a review

Iria Santos¹ · Luz Castro¹ · Nereida Rodriguez-Fernandez¹ · Álvaro Torrente-Patiño¹ · Adrián Carballal¹

Received: 22 May 2020 / Accepted: 1 December 2020 / Published online: 12 January 2021
© The Author(s), under exclusive licence to Springer-Verlag London Ltd. part of Springer Nature 2021

Abstract

In this article, we perform an exhaustive analysis of the use of Artificial Neural Networks and Deep Learning in the Visual Arts. We begin by introducing changes in Artificial Intelligence over the years and examine in depth the latest work carried out in prediction, classification, evaluation, generation, and identification through Artificial Neural Networks for the different Visual Arts. While we highlight the contributions of photography and pictorial art, there are also other uses for 3D modeling, including video games, architecture, and comics. The results of the investigations discussed show that the use of Artificial Neural Networks in the Visual Arts continues to evolve and have recently experienced significant growth. To complement the text, we include a glossary and table with information about the most commonly employed image datasets.

Keywords Artificial Neural Networks · Generative Adversarial Networks · Convolutional Neural Networks · Deep Learning · Visual Arts · Machine Learning · Prediction · Classification · Evaluation · Generation · Identification · Transfer Learning · Datasets

1 Introduction

Since the 1943 paper of McCulloch and Pitts [1], Artificial Neural Networks (ANNs) have been used for applications in many fields, such as health [2–4], optimization of structural design problems from civil engineering [5], traffic accident prediction [6], renewable energies [7], electrochemistry [8], video games generation [9], text translation [10], voice recognition [11–13], as well as in the applications of several types of commercialized hardware [14], etc.

Recent years have seen considerable growth in the use of Deep Learning, both in research and in industrial applications. Despite this recent growth, it has been used for decades, with examples including work from Giebel [15], Fukushima [16], and LeCun et al. [17]. But it was work like that of Krizhevsky et al. [18] that improved the state of the art of image classification. They trained a Deep

Convolutional Neural Network (DCNN) to classify 1.3 million high-resolution images in ImageNet LSVRC-2010 dataset [19] (1000 different classes). The tests achieved error rates of 39.7% (top 1) and 18.9% (top 5). To perform fast training, they used non-saturating neurons and a GPU to implement convolutional nets.

Techniques such as Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs) have been widely used. The basic architecture of a CNN consists of a convolution layer, a layer that applies the activation function in the matrix elements, a dimension reduction layer and a final layer with the number of neurons to be classified. This type of network can be made more complex, enhanced, and modified with other layers [20, 21]. GANs are another type of ANN in which two unsupervised networks of neurons compete. One of the networks (the generative network) generates candidates that the other ANN (the discriminative network) evaluates, following a scheme similar to that of co-evolutionary systems [22].

Few tasks are as characteristic of human beings as artistic ones. It is therefore natural that many scientists working in Artificial Intelligence (AI) are interested in modeling aspects of art using computer systems [23, 24].

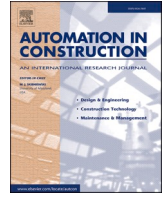
✉ Iria Santos
iria.santos@udc.es

¹ Department of Computer Science, Faculty of Computer Science, CITIC, University of A Coruña, 15071 A Coruña, Spain



Contents lists available at ScienceDirect

Automation in Construction

journal homepage: www.elsevier.com/locate/autcon

Review

Artificial intelligence applied to conceptual design. A review of its use in architecture

M. Luz Castro Pena^{*}, Adrián Carballal, Nereida Rodríguez-Fernández, Iria Santos, Juan Romero

Computer Science and Information Technology Department, CITIC-Research Center of Information and Communication Technologies, University of A Coruña, RNASA-IMEDIR Lab, ESCI. Campus de Elviña, 15008 A Coruña, Spain

ARTICLE INFO

Keywords:

Artificial intelligence
Architectural conceptual design
Evolutionary computing
Cellular automata

ABSTRACT

Conceptual architectural design is a complex process that draws on past experience and creativity to generate new designs. The application of artificial intelligence to this process should not be oriented toward finding a solution in a defined search space since the design requirements are not yet well defined in the conceptual stage. Instead, this process should be considered as an exploration of the requirements, as well as of possible solutions to meet those requirements.

This work offers a tour of major research projects that apply artificial intelligence solutions to architectural conceptual design. We examine several approaches, but most of the work focuses on the use of evolutionary computing to perform these tasks. We note a marked increase in the number of papers in recent years, especially since 2015. Most employ evolutionary computing techniques, including cellular automata. Most initial approaches were oriented toward finding innovative and creative forms, while the latest research focuses on optimizing architectural form.

1. Introduction

1.1. Conceptual design

As Song, Ghaboussi, and Kwon have indicated, architecture differs from other arts in that its products are required to be simultaneously aesthetically pleasing, structurally stable, and functional [1]. Determining a building's shape is the principal activity of the architectural design process. It is common for architects to start a design with a disembodied concept and a vague image of its shape which form the basis for proposing a broad set of solutions. The initial form will affect both performance and cost of construction, daylight use, energy consumption, layout configuration, shadow performance, acoustics, functional accessibility, and solar gain, among other features [2]. In this context, the search for shapes becomes one of the key steps in the conceptual design phase, as its results are inputs for the next steps in the design process, in the subsequent construction phase, and throughout the life cycle of the building.

Architectural design is a complicated process that draws on experience and creativity to develop new designs. Therefore, the application of artificial intelligence to this process should not be oriented to finding a

solution in a defined search space, since the design requirements are not yet well defined in the conceptual stage. Instead, this process should be considered as an exploration [3] of the requirements, as well as of possible solutions to meet those requirements [4,5]. Many design elements are chosen by considering a wide range of quantifiable and non-quantifiable features simultaneously. Even if a problem allows for numerical formulation, the lack of explicit and standard evaluation criteria makes defining design intentions difficult [6].

1.2. This review

In this work, we will examine the main research projects that applied artificial intelligence solutions to the design of form in architecture. As early as 1987, Soddu [7] created artificial DNA of Italian medieval cities which he used to define the Generative Design approach to Architecture and City Design in his book "Citta' Aleatorie." Since then, various approaches have been developed, such as Yeh's [8], which used an annealed neural network to find solutions to a facility layout problem applied to a hospital building case study in 2006. Wen, Hong, and Xueqiang in 2010 [9], and Rian and Asayama in 2016 [10] used fractal algorithms for the design of architectural forms. In related work,

^{*} Corresponding author at: RNASA-IMEDIR Lab, ESCI, Campus de Elviña, 15008 A Coruña, Spain
E-mail address: maria.luz.castro@udc.gal (M.L. Castro Pena).

<https://doi.org/10.1016/j.autcon.2021.103550>

Received 31 March 2020; Received in revised form 31 December 2020; Accepted 8 January 2021

Available online 22 January 2021







0926-5805/© 2021 The Authors.

Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Article

Validation of an Aesthetic Assessment System for Commercial Tasks

Nereida Rodriguez-Fernandez ^{†,‡}, Sara Alvarez-Gonzalez ^{†,‡}, Iria Santos ^{†,‡}, Alvaro Torrente-Patiño ^{†,‡},
Adrian Carballal ^{†,‡} and Juan Romero ^{*,†,‡}

Department of Computer Science and Information Technologies, Faculty of Communication Science, University of A Coruña, 15071 A Coruña, Spain; nereida.rodriguez@udc.es (N.R.-F.); sara.alvarezg@udc.es (S.A.-G.); iria.santos@udc.es (I.S.); alvaro.torrente@udc.es (A.T.-P.); adrian.carballal@udc.es (A.C.)

* Correspondence: jj@udc.es

† Current address: CITIC—Research Center of Information and Communication Technologies, University of A Coruña, 15071 A Coruña, Spain.

‡ These authors contributed equally to this work.

Abstract: Automatic prediction of the aesthetic value of images has received increasing attention in recent years. This is due, on the one hand, to the potential impact that predicting the aesthetic value has on practical applications. Even so, it remains a difficult task given the subjectivity and complexity of the problem. An image aesthetics assessment system was developed in recent years by our research group. In this work, its potential to be applied in commercial tasks is tested. With this objective, a set of three portals and three real estate agencies in Spain were taken as case studies. Images of their websites were taken to build the experimental dataset and a validation method was developed to test their original order with another proposed one according to their aesthetic value. So, in this new order, the images that have the high aesthetic score by the AI system will occupy the first positions of the portal. Relevant results were obtained, with an average increase of 52.54% in the number of clicks on the ads, in the experiment with Real Estate portals. A statistical analysis prove that there is a significant difference in the number of clicks after selecting the images with the AI system.

Keywords: aesthetics; e-commerce; real estate; validation; digital images



Citation: Rodriguez-Fernandez, N.; Alvarez-Gonzalez, S.; Santos, I.; Torrente-Patiño, A.; Carballal, A.; Romero, J. Validation of an Aesthetic Assessment System for Commercial Tasks. *Entropy* **2022**, *24*, 103. <https://doi.org/10.3390/e24010103>

Academic Editor: Amelia Carolina Sparavigna

Received: 24 October 2021

Accepted: 6 January 2022

Published: 9 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In a reality in which a large number of consumers are connected to the Internet, there are countless stimuli and options that they receive every day. On this basis, it is not surprising that more and more product choices are being made based on the aesthetic value and distinctiveness of visual design [1]. On the other side, companies are beginning to see the need to find new ways to be competitive, to attract and retain customers and to maintain or increase sales [2–4]. As Deng and Poole state, this requires a change in approach to design and user experience and a greater emphasis on aesthetics and the emotional responses it elicits. Through stimulated emotional responses, aesthetics can contribute to the success of e-commerce and encourage desirable user behaviours such as spending more time browsing, exploring more varied products, responding better to promotional initiatives, and increasing the likelihood of purchase [5]. In addition, after COVID-19, this adaptation must be even greater in the digital sphere, since the pandemic has contributed to the increase in the digitization of consumers [6]. In e-commerce, the products are perceived through their photographs and it is in them where consumers judge the aesthetic value discussed.

The great advances in technologies related to image capture and the growth of other fields such as visual computing [7,8] have given the average user more power and capacity in image acquisition. From the user's point of view, these new technologies create the

Article

Towards Automatic Image Enhancement with Genetic Programming and Machine Learning

João Correia ¹, Nereida Rodriguez-Fernandez ², Leonardo Vieira ¹, Juan Romero ^{2,*} and Penousal Machado ¹

¹ CISUC, Department of Informatics Engineering, University of Coimbra, 3030 Coimbra, Portugal; jncor@dei.uc.pt (J.C.); lmavieira@student.dei.uc.pt (L.V.); machado@dei.uc.pt (P.M.)

² Department of Computer Science and Information Technologies, Faculty of Communication Science, University of A Coruña, 15071 A Coruña, Spain; nereida.rodriguez@udc.es

* Correspondence:jj@udc.es

Abstract: Image Enhancement (IE) is an image processing procedure in which the image's original information is improved, highlighting specific features to ease post-processing analyses by a human or machine. State-of-the-art image enhancement pipelines apply solutions to fixed and static constraints to solve specific issues in isolation. In this work, an IE system for image marketing is proposed, more precisely, real estate marketing, where the objective is to enhance the commercial appeal of the images, while maintaining a level of realism and similarity with the original image. This work proposes a generic image enhancement pipeline that combines state-of-the-art image processing filters, Machine Learning methods, and Evolutionary approaches, such as Genetic Programming (GP), to create a dynamic framework for Image Enhancement. The GP-based system is trained to optimize 4 metrics: Neural Image Assessment (NIMA) technical and BRISQUE, which evaluate the technical quality of the images; and NIMA aesthetics and PhotoLike, that evaluate the commercial attractiveness. It is shown that the GP model was able to find the best image quality enhancement (0.97 NIMA Aesthetics), while maintaining a high level of similarity with the original images (Structural Similarity Index Measure (SSIM) of 0.88). The framework has better performance according to the image quality metrics than the off-the-shelf image enhancement tool and the framework's isolated parts.

Keywords: genetic programming; image enhancement; image filters; computer vision



Citation: Correia, J.; Rodriguez-Fernandez, N.; Vieira, L.; Romero, J.; Machado, P. Towards Automatic Image Enhancement with Genetic Programming and Machine Learning. *Appl. Sci.* **2022**, *12*, 2212. <https://doi.org/10.3390/app12042212>

Academic Editors: Mauro Castelli and Luca Manzoni

Received: 30 November 2021

Accepted: 29 January 2022

Published: 20 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

IE is an essential and ever-demanding branch of image processing and computer vision that allows visual improvement to images by manipulating their attributes. It can be used to alter an image in several different ways, for instance, by highlighting a specific feature to ease post-processing analyses by a human or a machine or by increasing its human perceived aesthetic. Each image has a set of attributes, such as size, color space, contrast, brightness, saturation, distortions, artifacts, noise, format, etc., that define how we, and different computer software solutions, perceive it. All these features are not isolated but, rather, interact with each other. The image format and compression, for instance, is pivotal to every other aspect of the image, as different formats allow for distinct color and compression properties that can, in turn, account for distinct visual characteristics. These attributes are often not well balanced between each other nor optimized for the image context and may cause a diverse range of deformations in the image quality.

Moreover, IE is useful for different tasks and fields that benefit from pre-processing an image to improve human or machine perception. Such fields are medical imagery, space imagery, bio-metric, photography, and video editing, among others [1–3]. Aesthetic improvement via IE is also one of the most popular uses of these techniques. A great example of such use of IE is the post-processing that is applied to an image before introducing it in a context where image appeal is crucial, for instance, in advertisements. Today, many consumers choose their purchases via the Internet, with countless stimuli and options.

Participación en congresos nacionales e internacionales

sciforum | EVENTS | ABOUT | Log In | NEW SUBMISSION | 🔍

Creative Computing and Computational Aesthetics

[Nereida Rodríguez-Fernandez](#)¹, [Iria Santos](#), [Alvaro Torrente-Patiño](#), [María Luz Castro Pena](#)

¹ RNASA-IMEDIR
² UDC

Academic Editor: Humbert G. Diaz

Published: 23 November 2021 by MDPI in *MOL2NET'21, Conference on Molecular, Biomedical & Computational Sciences and Engineering, 7th ed. congress AIMEDIC-08: AI, Med. Info., & Biomed. Eng. Congress, Coruña, Spain-Carleton, Canada-Stanford, USA, 2021*

<https://doi.org/10.3390/mol2net-07-11849>

Abstract:
Albert Szent-Györgyi defined a creative act as 'seeing what everyone else has seen and thinking as no one else has thought'. Would a machine be able to perform a creative act? Ada Lovelace, considered the first woman programmer, proposed almost 200 years ago the use of computers for artistic creative tasks, specifically musical ones. In recent years, our research group has been working on the creation of artificial artists and critics, as well as on classifiers and predictors of complexity and aesthetics. Automatic prediction of the aesthetic value of images has received increasing attention in recent years. This has come about, in part, because of the potential impact that aesthetic value has on practical applications. An aesthetic image evaluation system has been developed in recent years by our research group and has recently been tested for its potential and effectiveness in practical applications for commercial tasks. The results suggest an increase in the impact of advertisements using aesthetic criteria.

Keywords: creative computing; artificial artist; aesthetic value; image

[View paper](#)

[DOI](#) | [See at publisher website](#) | [Bookmark](#) | [Cite](#) | [Additional options](#) | [✉](#) | [f](#) | [t](#) | [in](#) | [s](#)



USANDO AMAZON MECHANICAL TURK PARA EVALUAR IMÁGENES

Nereida Rodríguez-Fernández ^{1,2}, Iria Santos ², Alvaro Torrente-Patiño ^{1,2}, María Luz Castro Pena ²

¹ RNASA-IMEDIR

² UDC

Published: 18 October 2020 by MDPI in MOL2NET'20, Conference on Molecular, Biomedical & Computational Sciences and Engineering, 6th ed. congress AIMEDIC-07: Int. Congress in Med. Info., UDC, Coruña, Spain-Sandford, USA, 2019

<https://doi.org/10.3390/mol2net-06-06961> (registering DOI)



Abstract:


En un mundo en el que cada vez tienen lugar más relaciones y transacciones a través de aplicaciones informáticas (compras, comunicaciones y relaciones personales), muchas decisiones se toman en función del valor estético, el atractivo o el impacto de imágenes digitales. Sin embargo, existen todavía pocos resultados sobre modelización estética; es decir, sobre la capacidad de realizar valoraciones y juicios estéticos.

Diferentes grupos de investigadores han intentado crear sistemas informáticos capaces de aprender la percepción estética de un grupo de seres humanos como parte de un sistema generativo, con la intención de ser utilizados en la selección u ordenación automática de imágenes. Dada la naturaleza subjetiva del problema estético, la selección del conjunto de datos con el que se entrena el sistema es especialmente relevante.

Tras haber analizado el grado de generalización de algunos conjuntos muestrales utilizados para la clasificación automática de imágenes, se ha concluido que no es suficiente para tomarlos como referencia en el entrenamiento de sistemas de predicción y clasificación automática de imágenes. Además, se han detectado también otras limitaciones funcionales en dichos datasets.


Con la intención de ofrecer una solución a los problemas detectados, se presenta la creación de un nuevo conjunto de imágenes procedentes del portal web DPChallenge.com, con mayor coherencia estadística. Este nuevo dataset ha sido evaluado según criterios de estética y de calidad por un grupo de individuos españoles en condiciones experimentales controladas y por otro grupo de estadounidenses a través de encuestas online. Así, este dataset se convierte en el primer conjunto de imágenes evaluado por tres poblaciones diferentes (el que evalúa en el portal web DPChallenge.com, el español y el

Springer Link Search   Log in



International Conference on Computational Intelligence in Music, Sound, Art and Design (Part of EvoStar)
 ↳ EvoMUSART 2021: **Artificial Intelligence in Music, Sound, Art and Design** pp 82–97 | [Cite as](#)

Evolving Image Enhancement Pipelines

João Correia , Leonardo Vieira, Nereida Rodriguez-Fernandez, Juan Romero & Penousal Machado

Conference paper | [First Online: 02 April 2021](#)

1045 Accesses | [2 Citations](#)

Part of the [Lecture Notes in Computer Science](#) book series ([LNTCS, volume 12693](#))

Abstract

Image enhancement is an image processing procedure in which the original information of the image is improved. It alters an image in several different ways, for instance, by highlighting a specific feature in order to ease post-processing analyses by a human or machine. In this work, we show our approach to image enhancement for digital real-estate-marketing. The aesthetic quality of the images for real-estate marketing is critical since it is the only input clients have once browsing for options. Thus, improving and ensuring the aesthetic quality of the images is crucial for marketing success. The problem is that each set of images, even for the same real-estate item, is often taken under diverse conditions making it hard to find one solution that fits all. State of the art image enhancement pipelines applies a set of filters that solve specific

Access via your institution →

Chapter EUR 29.95

Price includes VAT (Spain)

- DOI: 10.1007/978-3-030-72914-1_6
- Chapter length: 16 pages
- Instant PDF download
- Readable on all devices
- Own it forever
- Exclusive offer for individuals only
- Tax calculation will be finalised during checkout

Buy Chapter

eBook EUR 71.68

Softcover Book EUR 88.39

[Learn about institutional subscriptions](#)

Sections

Figures

References

Dataset for the Aesthetic Value Automatic Prediction [†]

Nereida Rodriguez-Fernandez *, Iria Santos and Alvaro Torrente

Department of Computer Science, Faculty of Computer Science, University of A Coruña, 15071 A Coruña, Spain; iria.santos@udc.es (I.S.); alvaro.torrente@udc.es (A.T.)

* Correspondence: nereida.rodriguez@udc.es

† Presented at 2nd XoveTIC Conference, A Coruña, Spain, 5–6 September 2019.

Published: 1 August 2019

Abstract: One of the most relevant issue in the prediction and classification of the aesthetic value of an image is the sample set used to train and validate the computational system. In this document the limitations found in different datasets used to classificate and predict aesthetic values are exposed, and a new dataset is proposed with images from the DPChallenge.com portal, with evaluations of three different populations.

Keywords: dataset; aesthetics; quality; prediction; classification; artificial intelligence; assessment

1. Introduction

Different research groups have tried to create computer systems capable of learning the aesthetic perception of a group of human beings as part of a generative system, with the intention of being used in the selection or automatic ordering of images. Due the subjective nature of the aesthetic problem, the selection of the dataset with which the system is trained is especially relevant. After analyzed, in previous research [1,2], the generalization degree of some datasets, it has been concluded that it is not enough to take them as a reference in the training of automatic image classification and prediction systems. In order to providing a solution to the problems detected, this paper describes the creation of a new dataset from the DPChallenge.com portal, with greater statistical coherence. In addition, this new dataset has been evaluated according to aesthetic and quality criteria by a human group in controlled experimental conditions and by another American group through online surveys.

2. Limitations Found in the Datasets Available

There are some datasets that have been used in several times for the images classification. Among them, Photo.net [3–5], DPChallenge.com [6,7] and the one created by Cela-Conde et al. [8–10] However, when its generalization capacity is studied, it has been detected that they cannot be considered as representative for the realization of image experiments. In some cases, the correlation is greater when the validation set belongs to the same data source as the training set, and this correlation drops markedly when the validation source set is different from that of the training. In addition, the sample sets trained with evaluations from the photographic portals have some defects: the evaluation system does not have the same control as a psychological test because it is not possible to obtain all the information about the evaluators or about the evaluation conditions; the number of images could be insufficient, since there is no justified reason to choose a sample size and there is a very high difference between the number of people who value each image; user ratings can be easily conditioned by personal tastes, personal relationships with the work creator, or by the momentary boom or popularity of certain styles. Lastly, in one of the cases [3] it has been shown that the users of these portals do not have sufficient grounds to differentiate between aesthetics and

Digital Image Quality Prediction System [†]

Nereida Rodriguez-Fernandez ^{1,2,*} , Iria Santos ^{1,2} , Alvaro Torrente-Patiño ³ 
and Adrian Carballal ^{1,3} 

¹ CITIC-Research Center of Information and Communication Technologies, University of A Coruña, 15071 A Coruña, Spain; iria.santos@udc.es (I.S.); adrian.carballal@udc.es (A.C.)

² Department of Computer Science and Information Technologies, Faculty of Communication Science, University of A Coruña, Campus Elviña s/n, 15071 A Coruña, Spain

³ Department of Computer Science and Information Technologies, Faculty of Computer Science, University of A Coruña, Campus Elviña s/n, 15071 A Coruña, Spain; alvaro.torrente@udc.es

* Correspondence: nereida.rodriguez@udc.es

[†] Presented at the 3rd XoveTIC Conference, A Coruña, Spain, 8–9 October 2020.

Published: 19 August 2020



Abstract: “A picture is worth a thousand words.” Based on this well-known adage, we can say that images are important in our society, and increasingly so. Currently, the Internet is the main channel of socialization and marketing, where we seek to communicate in the most efficient way possible. People receive a large amount of information daily and that is where the need to attract attention with quality content and good presentation arises. Social networks, for example, are becoming more visual every day. Only on Facebook can you see that the success of a publication increases up to 180% if it is accompanied by an image. That is why it is not surprising that platforms such as Pinterest and Instagram have grown so much, and have positioned themselves thanks to their power to communicate with images. In a world where more and more relationships and transactions are made through computer applications, many decisions are made based on the quality, aesthetic value or impact of digital images. In the present work, a quality prediction system for digital images was developed, trained from the quality perception of a group of humans.

Keywords: machine learning; genetic algorithm; quality; image; prediction; dataset

1. Introduction

In recent years, significant efforts were applied to the development of successful models and algorithms that can automatically and accurately predict the perceptual quality of two-dimensional (2D) and three-dimensional (3D) digital images and videos. This estimate comes from studies with at least a century of experience, or more if we take into account those developed by Platon and Aristotle, usually from Humanities departments: Psychology, Sociology, Philosophy, Fine Arts, etc. [1]. Different research groups sought to create computer systems capable of learning the aesthetic and quality perception of a group of humans as part of a generative system for uses such as the selection and arrangement of images within a set, even though it is complex to translate this into computer problems. Visual quality refers to the quantification of the perceptual degradation of a visual stimulus due to the presence or absence of distortions. Most of the applications that were developed were designed to treat synthetically distorted images [2]. In this case, unlike other image quality assessment algorithms that use synthetically distorted images [3,4], it was decided to use images with absence of distortion [5,6]. Despite the fact that the data collected contained quality and aesthetic results, on this occasion only the quality data were used as they constituted more objective results [7].

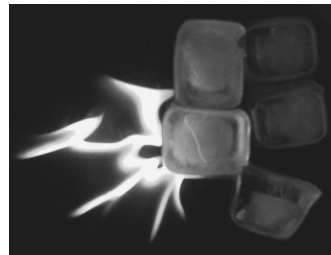
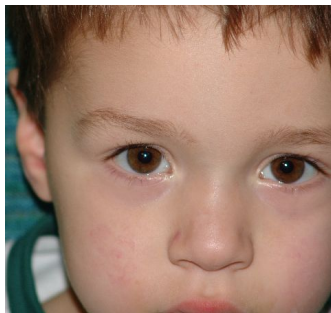
Apéndice B

Galería de imágenes del dataset propuesto en el capítulo 5

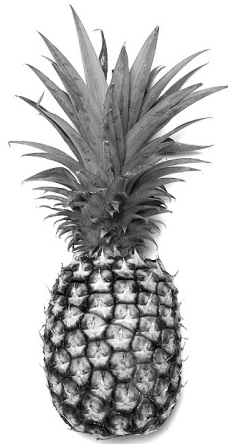
En este apéndice se presenta una muestra de las imágenes que componen el dataset propuesto en el capítulo 5. En cada sección se muestran ordenadas por rangos en función de cada una de las puntuaciones que forman el dataset (DPChallenge, calidad y estética). Es necesario recordar algunas cuestiones sobre este dataset. En primer lugar, las puntuaciones de DPChallenge se encuentran en un rango de puntuación de entre 1 y 10, mientras que las de calidad y estética se puntuaron en un rango de entre 1 y 5. Por otra parte, para la construcción de un conjunto de datos con una representación coherente de todos los rangos de puntuación, en el conjunto de DPChallenge, se eliminaron las imágenes con una puntuación media inferior a 3 y superior a 8, por no alcanzar la cuenta mínima de 200 que se puso como condición.

Imágenes ordenadas según la puntuación de DP-Challenge

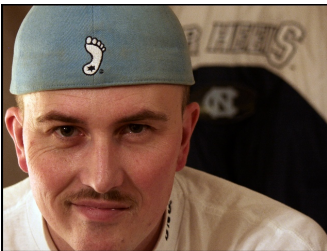
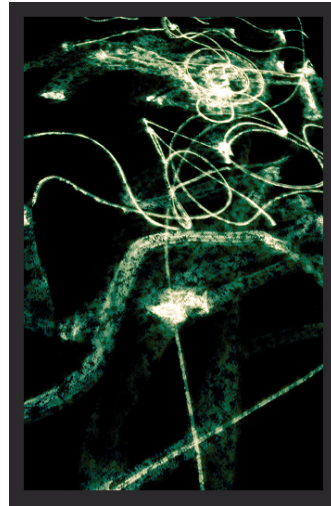
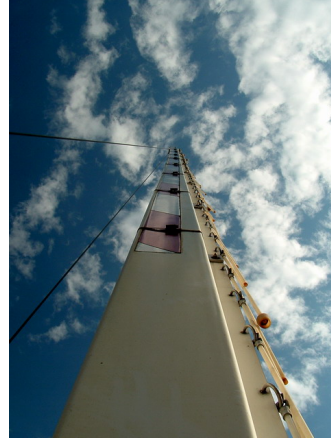
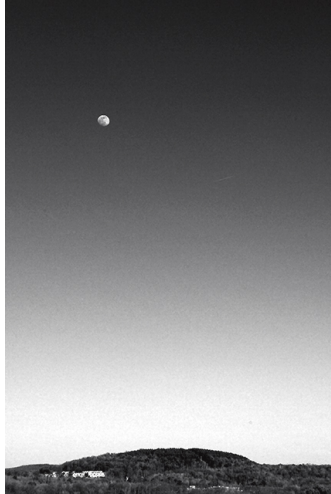
Ejemplos de imágenes del rango [3,4)



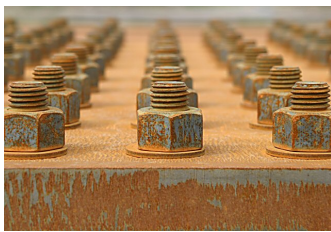
Ejemplos de imágenes del rango [4,5)



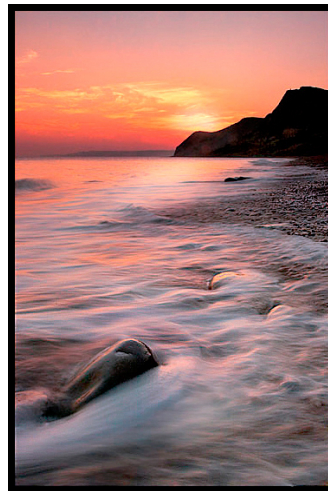
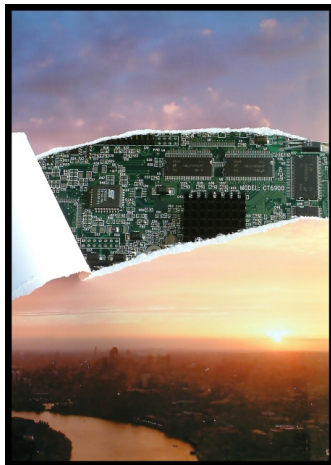
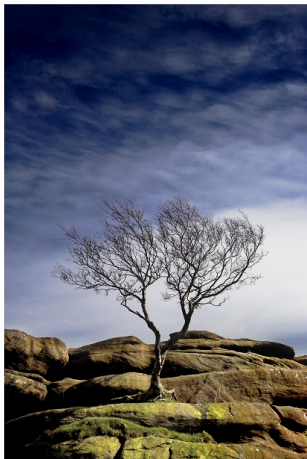
Ejemplos de imágenes del rango [5,6)



Ejemplos de imágenes del rango [6,7)

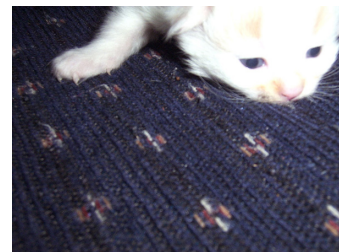


Ejemplos de imágenes del rango [7,8)

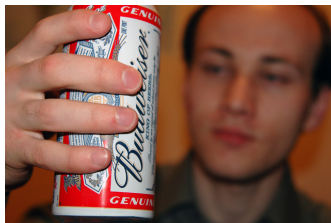
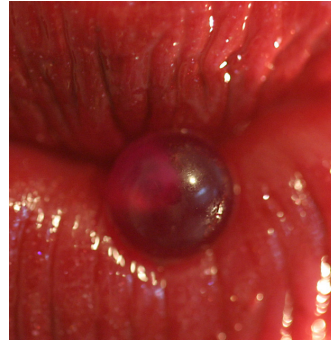
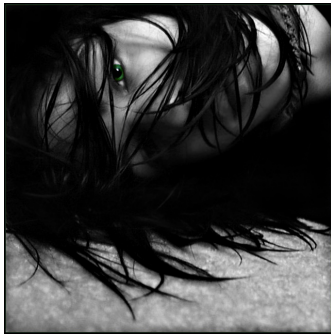


Imágenes ordenadas según su valor de calidad

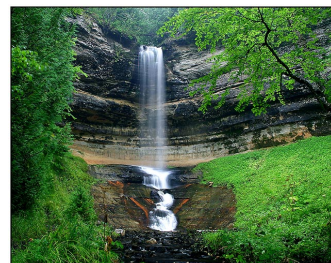
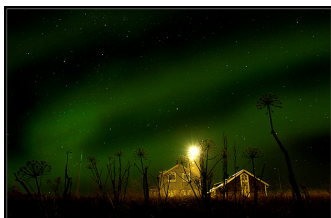
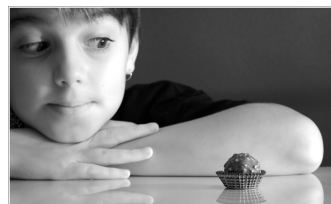
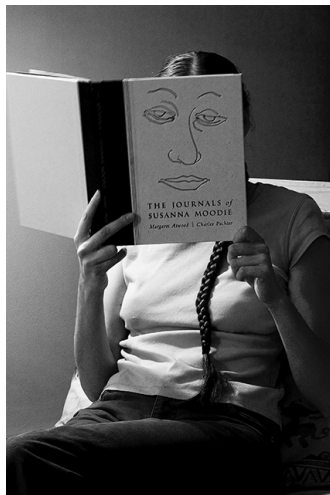
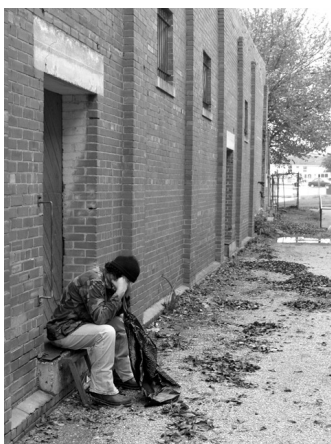
Ejemplos de imágenes del rango [1,2)



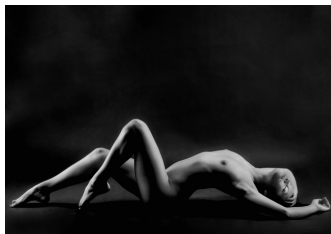
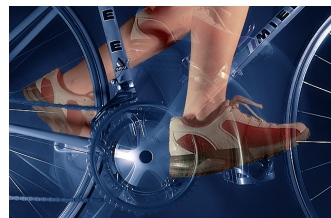
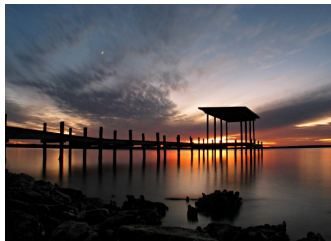
Ejemplos de imágenes del rango [2,3)



Ejemplos de imágenes del rango [3,4)

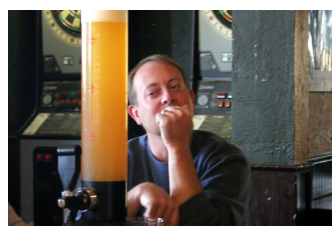
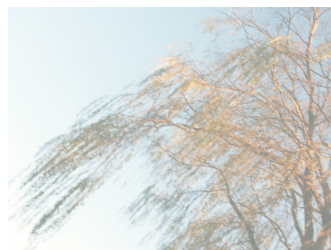
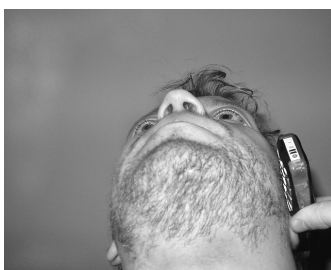
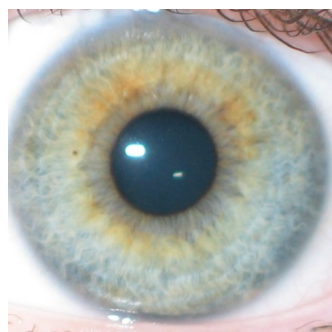


Ejemplos de imágenes del rango [4,5)

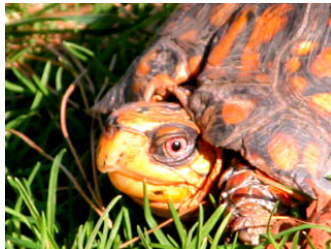
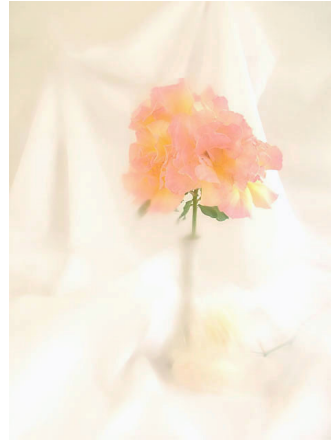


Imágenes ordenadas según su valor estético

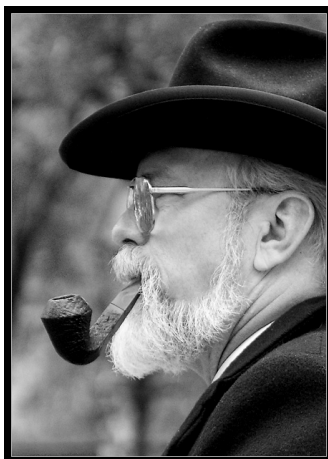
Ejemplos de imágenes del rango [1,2)



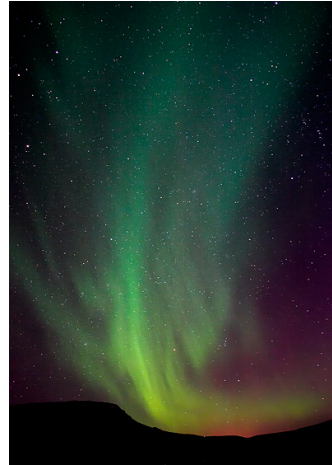
Ejemplos de imágenes del rango [2,3)



Ejemplos de imágenes del rango [3,4)



Ejemplos de imágenes del rango [4,5)


















Apéndice C

Galería de imágenes utilizadas en el capítulo 7

En esta sección se exponen de forma comparativa las 10 primeras imágenes de cada conjunto utilizado en la fase experimental del capítulo 7, presentadas de mayor a menor según su orden original, según su valor estético y según el número de votos recibidos por los participantes en el proceso de evaluación.

Imágenes del Portal1

	Original	Estético	Nº de votos
1			
2			
3			
4			
5			

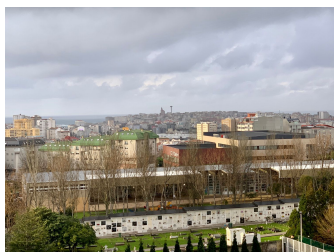
6



7



8


















9



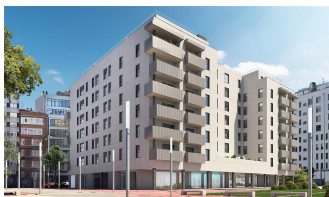
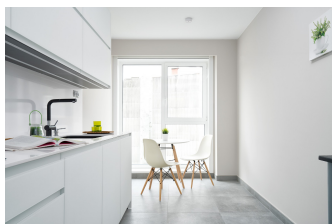
10



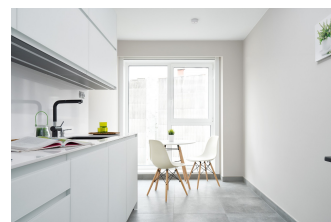
Imágenes del Portal2

	Original	Estético	Nº de votos
1			
2			
3			
4			
5			

6



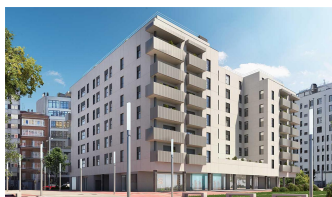
7



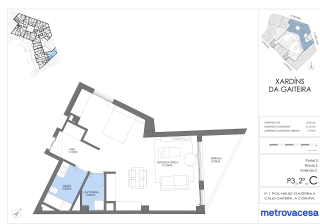
8


















9



10



Imágenes del Portal3

	Original	Estético	Nº de votos
1			
2			
3			
4			
5			

6



7



8















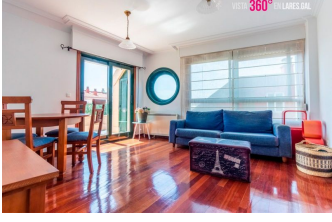


9

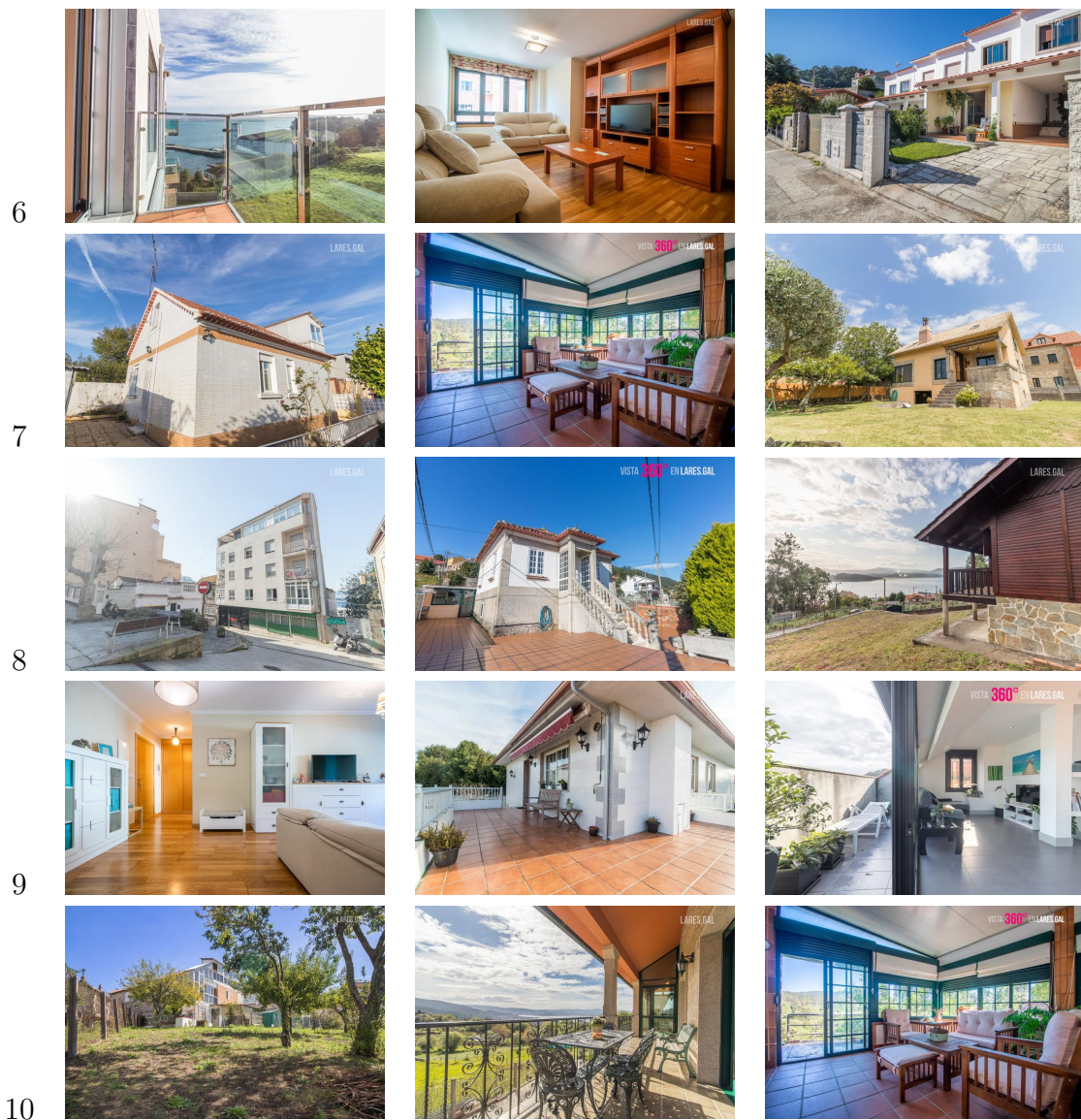


10


















Imágenes de la Agencial

	Original	Estético	Nº de votos
1			
2			
3			
4			
5			



Imágenes de la Agencia2

	Original	Estético	Nº de votos
1			
2			
3			
4			
5			

6



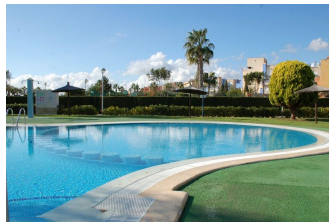
7



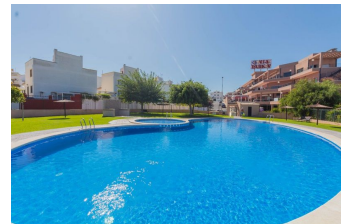
8



9



10



Imágenes de la Agencia3

	Original	Estético	Nº de votos
1			
2			
3			
4			
5			

