ADDRESSING CLASS IMBALANCE IN MULTI-CLASS IMAGE CLASSIFICATION BY MEANS OF AUXILIARY FEATURE SPACE RESTRICTIONS

M. Dorozynski[1,*], F. Rottensteiner[1]

[1] Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover, Germany
(dorozynski, rottensteiner)@ipi.uni-hannover.de

Commission II, WG II/8

KEY WORDS: Deep learning, convolutional neural networks, image classification, class imbalances, auxiliary clustering loss, silk heritage.

ABSTRACT:

Learning from imbalanced class distributions generally leads to a classifier that is not able to distinguish classes with few training examples from the other classes. In the context of cultural heritage, addressing this problem becomes important when existing digital online collections consisting of images depicting artifacts and assigned semantic annotations shall be completed automatically; images with known annotations can be used to train a classifier that predicts missing information, where training data is often highly imbalanced. In the present paper, combining a classification loss with an auxiliary clustering loss is proposed to improve the classification performance particularly for underrepresented classes, where additionally different sampling strategies are applied. The proposed auxiliary loss aims to cluster feature vectors with respect to the semantic annotations as well as to visual properties of the images to be classified and thus, is supposed to help the classifier in distinguishing individual classes. We conduct an ablation study on a dataset consisting of images depicting silk fabrics coming along with annotations for different silk-related classification tasks. Experimental results show improvements of up to 10.5% in average F1-score and up to 20.8% in the F1-score averaged over the underrepresented classes in some classification tasks.

1. INTRODUCTION

The EU H2020 project SILKNOW (`https://silknow.eu/`) was dedicated to the preservation and better understanding of the European silk heritage. In this context, a database of silk artifacts based on publicly available online collections has been generated, where for each record (each artifact), meta-information, e.g. information related to the time or place of production, is available in a standardized way. However, this information is not always readily available in the original collections. For such records, the relevant properties of silk fabrics (e.g. the *manufacturing technique* or the *material*) are to be predicted automatically from images of the artifacts. In (Dorozynski et al., 2019), this task was addressed by training a convolutional neural network (CNN; (LeCun et al., 1989; Krizhevsky et al., 2012)) that takes the image of a fabric as input and predicts the corresponding class labels. The CNN was trained using labelled training images, i.e. images for which the true labels are known in advance for the relevant properties. Whereas reasonable overall accuracies could be achieved, we could also observe that the class-specific F1 scores varied considerably, mainly depending on the number of available training samples per class. We attribute these problems to a strong *class imbalance* of the training data, resulting in a classifier that performs very well for dominant classes having many samples, while delivering worse results for the underrepresented classes, i.e. the classes having few training samples. For integrating image-based predictions in the silk database, it is of special interest to apply a classifier that is able to distinguish the classes of all silk properties well such that added value is delivered for the user of the silk database through the predictions.

It is a well-known problem that training using data with imbalanced class distributions results in a classifier that tends to predict classes that were well represented in training data rather well, whereas classes with only few examples in training data often cannot be distinguished from other classes (Krawczyk, 2016; Johnson and Khoshgoftaar, 2019; Sridhar and Kalaivani, 2021). Early approaches addressing class imbalance problems proposed to artificially balance the class distributions by oversampling of classes with few examples, e.g. (Chawla et al., 2002), or by undersampling of classes with many examples, e.g. (Mani and Zhang, 2003). Whereas sampling methods are also investigated for learning classifiers based on CNNs, e.g. (Pouyanfar et al., 2018), learning image features by CNNs opens up new possibilities for dealing with imbalanced training data. Using margin constraints in the loss function concerning differences between the feature vectors to be learned, features of examples belonging to the same class can be forced to be close together in feature space and features related to different classes can be forced to be further away, e.g. (Huang et al., 2016; Hameed et al., 2021). Thus, the feature vectors are clustered such that each cluster in feature space belongs to one class of the classification problem. Nevertheless, these approaches come along with further training hyper-parameters, i.e. the distance margins in (Huang et al., 2016) or the angular margins in (Hameed et al., 2021). Additionally, the clustering exclusively relies on semantic aspects, because the clustering criterion is exclusively based on the class labels of the images that are represented by the features.

To the best of our knowledge, addressing the problem of class imbalances in the context of cultural heritage datasets has not been investigated. Like preceding work, we will address class imbalance problems by a clustering of feature vectors produced by a CNN, but in the context of silk heritage. In contrast to existing work, the features are not only clustered with respect to

* Corresponding author

semantic information, but also with respect to visual properties of the input images. Thus, our scientific contributions are the following:

- We propose a new auxiliary clustering loss for handling class imbalances in a multi-class classification scenario.
- In this context, clustering is performed with respect to visual and semantic properties of the images to be classified.
- Furthermore, the clustering loss supports both inter-class separability and intra-class connectivity without the need for additional training hyper parameters.
- Finally, comprehensive experiments are conducted to investigate the impact of the components of the auxiliary loss on the classification performance for different silk-related classification tasks, each with a different number of classes and a different degree of class imbalance.

2. RELATED WORK

Learning from imbalanced training data is a well known problem in the domains of Photogrammetry and Computer Vision (Johnson and Khoshgoftaar, 2019; Sridhar and Kalaivani, 2021). In the context of learning using data with imbalanced class distributions, the resulting classifiers tend to show a weak performance in correctly predicting examples from classes with few training data, which is a challenge both in binary and multi-class classification (Krawczyk, 2016). Different strategies have been applied to deal with this problem, where the corresponding methods can be categorized as data-level methods, algorithmic-level methods and hybrid methods (Krawczyk, 2016; Johnson and Khoshgoftaar, 2019). Data-level methods aim to compensate imbalances in the training data by oversampling of classes with few examples, e.g. (Chawla et al., 2002), by undersampling classes with many examples, e.g. (Mani and Zhang, 2003), or by performance-driven dynamic sampling in each training step, e.g. (Pouyanfar et al., 2018). Algorithmic-level methods such as (Ling and Sheng, 2008; Lin et al., 2017) adapt the training objectives such that classes with few training examples have a higher impact on the classifier's parameters, and hybrid methods, e.g. (Dong et al., 2018), combine aspects of both data-level methods and algorithmic-level methods. In contrast to approaches aiming to increase the impact of examples belonging to underrepresented classes on determining the classifier's parameters during training or to carefully select representative training examples, we focus on an adequate separation of the classes in feature space, which we believe to be helpful for distinguishing all classes.

According to Krawczyk (2016), class imbalance may be irrelevant if there are sufficient representations for both, frequent as well as less frequent classes. Using CNNs (LeCun et al., 1989; Krizhevsky et al., 2012), representations of images to be used for classification can be learned effectively. Thus, one way of achieving such a sufficient representation is to guide the CNN to learn that the feature vectors belonging to the same class should form a distinct cluster in feature space and that clusters corresponding to different classes should be different from each other. Thus, combining classification and clustering in training could help to mitigate the problems related to class imbalance of the training data. Existing work that combines image classification and clustering in feature space exploits k-means clustering to obtain pseudo-labels for learning a classifier, e.g. (Caron et al., 2018; Yang et al., 2021; Ma et al., 2021). There is further work that exploits clustering as auxiliary training constraint for learning a classifier. The basic principle is to combine a classification loss with an auxiliary metric learning loss. Wen et al. (2016) aim to support intra-class connectivity by forcing all feature vectors related to one class to be close to the corresponding center of the feature vectors using an auxiliary center loss. Qi and Su (2017) expand the center loss by an additional term such that it also requires inter-class separability. Instead of forcing the distances in feature space to be small for features belonging to the same class and large for features belonging to different classes, respectively (Wen et al., 2016; Qi and Su, 2017), there are also margin-based loss variants that introduce within-class and between-class margins to explicitly force the produced clusters to reflect inter-class separability and intra-class connectivity. Whereas distance-based margin constraints are proposed in (Huang et al., 2016; Liu et al., 2017; Yang et al., 2020) , the approaches in (Choi et al., 2020; Hameed et al., 2021) rely on angular margins. However, margin-based losses require at least one further hyper-parameter defining the appropriate cluster size; it would be desirable not having to tune such a parameter.

Even though all work mentioned so far addressed class imbalance problems, none of them focuses on handling imbalanced distributions in the context of cultural heritage applications. Training image-based classifier to predict semantic information of historically relevant artifacts on the basis of images of the same is a growing field of research. Training a classifier to predict painting properties such as the artist, the genre or the style was investigated in several works, e.g. (Blessing and Wen, 2010; Tan et al., 2016; Sur and Blaine, 2017). Whereas Blessing and Wen (2010) learn a support vector machine for the prediction of a painting's artist, a CNN-based classifier for predicting the artist is trained in (Tan et al., 2016; Sur and Blaine, 2017); Tan et al. (2016) also deals with the prediction of the style and the genre of a painting. Instead of training one classifier per relevant classification task, multiple tasks can be jointly solved in the frame of multi-task learning (MTL) to improve generalization of the resulting classifier (Caruana, 1993). Even though all classification losses are equally important in MTL, it can also be seen as learning each classification task with auxiliary classification losses, which has also been applied in the domain of cultural heritage-related image classification, e.g. (Strezoski and Worring, 2017; Bianco et al., 2019; Garcia et al., 2020). No approaches were found that investigate training a classifier with auxiliary losses aiming to predict historically relevant information on the basis of images. In particular, to the best of our knowledge, there is no paper that investigates class imbalance problems in this context.

Accordingly, this is the first work in the context of cultural heritage that tries to tackle class imbalance problems by exploiting an auxiliary clustering loss. Even though training a classifier while forcing feature vectors to form class-related clusters was investigated in other contexts, e.g. (Liu et al., 2017; Choi et al., 2020), we do not require additional hyper-parameter tuning for cluster definitions. The approaches in (Wen et al., 2016; Qi and Su, 2017) also do not introduce additional hyper-parameters in the auxiliary clustering loss, but in contrast to those works as well as the works investigating margin-based approaches, the features in our work are not only clustered based on the available class information but also based on the visual properties of the related input images. Nevertheless, the proposed clustering

loss forces the distances of the feature vectors to reflect intra-class connectivity and inter-class separability. The most similar work to the present one is our earlier work in (Dorozynski and Rottensteiner, 2022) dealing with descriptor learning for image retrieval exploiting an auxiliary classification loss. Even though the loss formulation is similar, the focus of the present work is image classification, while descriptor learning is applied to support a good clustering of the feature vectors.

3. METHODOLOGY

In this paper, we propose a training strategy for a CNN-based image classifier that combines classification and feature space clustering. During training, the classification loss is jointly minimized with an auxiliary clustering loss. The goal of the clustering loss is to support classification by producing an appropriate image representation. For clustering, we exploit similarity losses proposed in the context of descriptor learning in (Dorozynski and Rottensteiner, 2022) aiming to support intra-class compactness as well as inter-class separability. We assume that feature vectors that form clusters in feature space so that each cluster belongs to a different class will help a classifier to distinguish the classes to be learned and, thus, also to correctly predict the labels of samples belonging to underrepresented classes. Even though we train our classifier with auxiliary losses, no additional input data is needed; the proposed training strategy requires images with assigned class labels for the classification task to be learned both for the classification loss and the auxiliary clustering losses. Section 3.1 will contain a description of the network architecture and section 3.2 gives details about the training strategy.

3.1 Network architecture

The proposed classifier takes an RGB image x as an input and delivers normalized class scores $y_k(x)$ for all K classes to be distinguished in a classification task as depicted in figure 1. First of all, the image x is mapped to a 2048-dimensional feature vector $f_{RN}(x)$ by means of a ResNet152 backbone (He et al., 2016) with parameters $\mathbf{w}_{RN}$, followed by a ReLU activation (rectified linear unit (Nair and Hinton, 2010)), a dropout layer (Srivastava et al., 2014) with a 30% dropout rate, and a sub-network *task fc* consisting of NL fully connected layers with $NN^1, ..., NN^{NL}$ nodes, resulting in the feature vector $f_{tfc}(x)$. The parameters of this sub-network are denoted by $\mathbf{w}_{tfc}$. Afterwards, the vector $f_{tfc}(x)$ is presented to both a classification head as well as to a clustering head, where the clustering head is only active during training. The classification head starts with a ReLU activation applied to $f_{tfc}(x)$ and maps the feature vectors to normalized class scores $y_k(x), k = 1, ..., K$ by means of a softmax layer with parameters $\mathbf{w}_{sm}$. The clustering head takes the vector $f_{tfc}(x)$ and normalizes it to unit length, resulting in a vector $f_{aux}(x)$.

3.2 Network training

Training of a CNN is realized by iteratively updating the network weights such that a loss function is minimized. For learning a classifier with an auxiliary clustering loss, the CNN in Figure 1 can be trained by minimizing the loss function

$$\mathcal{L}(\mathbf{x}, \mathbf{w}) = \lambda_C \cdot \mathcal{L}_C(\mathbf{x}, \mathbf{w}) + \lambda_{aux} \cdot \mathcal{L}_{aux}(\mathbf{x}, \mathbf{w}_v) \quad (1)$$

for an image set $\mathbf{x}$, where $\mathbf{w}_v := [\mathbf{w}_{RN}^T, \mathbf{w}_{tfc}^T]^T$ denotes the set of weights affecting the clustering head of the network and

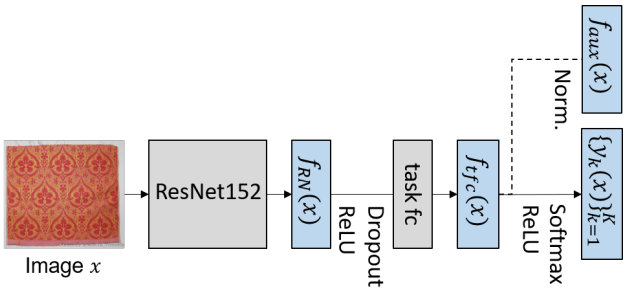

Figure 1. CNN architecture. An input image x is presented to a ResNet152 (He et al., 2016), resulting in a feature vector $f_{RN}(x)$, which is then mapped to a task-specific representation $f_{tfc}(x)$ by a sub-network *task fc*. $f_{tfc}(x)$ is presented both to a classification head and to a clustering head. The classification head takes $f_{tfc}(x)$ as input for a ReLU activation and maps the output to normalized class scores $y_k(x)$ of every class k using a softmax layer. The clustering head consists of a normalization of $f_{tfc}(x)$ to unit length, resulting in a feature vector $f_{aux}(x)$. The broken line indicates that the clustering head is only active in training.

$\mathbf{w} := [\mathbf{w}_v^T, \mathbf{w}_{sm}^T]^T$ is a vector containing all network weights. The total loss $\mathcal{L}(\mathbf{x}, \mathbf{w})$ is a weighted sum of the classification loss $\mathcal{L}_C(\mathbf{x}, \mathbf{w})$ and the proposed auxiliary clustering loss $\mathcal{L}_{aux}(\mathbf{x}, \mathbf{w}_v)$, with the parameters $\lambda_C, \lambda_{aux} \in [0, 1]$ controlling the influence of the individual loss terms. The classification loss is chosen to be the standard softmax cross entropy loss (Bishop, 2006) aiming to assign high normalized class scores to the true classes.

The auxiliary clustering loss $\mathcal{L}_{aux}(\mathbf{x}, \mathbf{w}_v)$ is supposed to adapt the network weights $\mathbf{w}_v$ such that the feature vectors of images belonging to the same class are forced to be close together in feature space, leading to intra-class connectivity, whereas feature vectors of images belonging to different classes are forced to be far away in feature space, leading to inter-class separability. We assume that images belonging to the same class are both semantically and visually similar in some respect, whereas images belonging to different classes tend to be dissimilar with respect to their semantic and visual properties. Thus, determining the network weights such that the Euclidean distance of feature vectors reflects the degree of similarity of the respective images is supposed to deliver the desired clustering. For that purpose, the clustering loss

$$\mathcal{L}_{aux}(\mathbf{x}, \mathbf{w}_v) = \alpha_{sem} \cdot \mathcal{L}_{sem}(\mathbf{t}, \mathbf{w}_v) + \alpha_{co} \cdot \mathcal{L}_{co}(\mathbf{p}, \mathbf{w}_v), \quad (2)$$

consisting of a semantic similarity loss $\mathcal{L}_{sem}(\mathbf{t}, \mathbf{w}_v)$ taking triplets of images $\mathbf{t}$ as input and a colour similarity loss $\mathcal{L}_{co}(\mathbf{p}, \mathbf{w}_v)$ taking image pairs $\mathbf{p}$ as input is proposed. The two parameters $\alpha_{sem}, \alpha_{co} \in [0, 1]$ in equation 2 control the influence of the two similarity losses. Both similarity losses were proposed in (Dorozynski and Rottensteiner, 2022), where the term $\mathcal{L}_{sem}$ is designed to consider semantic similarity and the term $\mathcal{L}_{co}$ considers colour similarity.

According to Dorozynski and Rottensteiner (2022), images having similar semantic properties, i.e. similar class labels, are considered to be semantically similar, and images with dissimilar semantic properties are considered to be dissimilar. As we train a separate classifier for all variables, we use a special case of the semantic similarity loss in (Dorozynski and Rottensteiner, 2022), defining that loss only on the basis of $M = 1$

semantic variable. Furthermore, in contrast to the original approach, all images used to train a classifier come along with a class label for the task to be learned. Thus, the semantic similarity loss can be simplified, leading to

$$\mathcal{L}_{sem}(\mathbf{t}, \mathbf{w}_v) = \frac{1}{N_t} \cdot \sum_{n_t=1}^{N_t} max\left(M_{i,p,n}^{n_t} + \Delta_{i,p,\mathbf{w}_v}^{n_t} - \Delta_{i,n,\mathbf{w}_v}^{n_t}, 0\right)$$

$$M_{i,p,n}^{n_t} = \delta_{i,p} - \delta_{i,n} \overset{!}{>} 0. \tag{3}$$

The semantic similarity loss $\mathcal{L}_{sem}(\mathbf{t}, \mathbf{w}_v)$ is calculated for all N_t triplets $\mathbf{t}$ in the image set $\mathbf{x}$, where a triplet consists of an anchor sample x_i, a positive sample x_p belonging to the same class as x_i and a negative sample x_n belonging to a different class than the pair (x_i, x_p). The loss forces the Euclidean distance $\Delta_{i,n,\mathbf{w}_v}^{n_t}$ between the feature vectors $f_{aux}(x_i), f_{aux}(x_n)$ of (x_i, x_n) to be larger than the distance $\Delta_{i,p,\mathbf{w}_v}^{n_t}$ between the feature vectors $f_{aux}(x_i), f_{aux}(x_p)$ of (x_i, x_p) by at least a margin $M_{i,p,n}^{n_t}$. The margin in equation 3, where $\delta_{i,o} = 1$ in case (x_i, x_o) belong to the same class and $\delta_{i,o} = 0$ in all other cases, requires x_p to be more similar to x_i than x_n indeed. The inequality requiring the margin to be larger than zero is used as a constraint to select valid triplets. Accordingly, the semantic similarity loss is only calculated for triplets fulfilling this margin constraint, i.e. for the N_t valid triplets $x_i, x_p, x_n \in \mathbf{x}$ fulfilling the inequality constraint in equation 3.

Similarly, the colour similarity loss considers visual aspects of the images and forces the distances between pairs of feature vectors to correspond to the similarity of the colour distributions in HSV colour space of the respective images. The similarity of the colour distributions is expressed by the correlation coefficient $\rho\,(x_i, x_o) \in [-1; 1]$ (Dorozynski and Rottensteiner, 2022), where $\rho\,(x_i, x_o) = 1$ indicates 100% colour similarity of x_i, x_o and lower values of $\rho\,(x_i, x_o)$ indicate a lower degree of colour similarity. The colour similarity loss

$$\mathcal{L}_{co}(\mathbf{p}, \mathbf{w}_v) = \frac{1}{N_{co}} \cdot \sum_{n_{co}=1}^{N_{co}} max\left(0, |\Delta_{i,o,\mathbf{w}_v}^{n_{co}} - M_{i,o}^{n_{co}}|\right)$$

$$M_{i,o}^{n_{co}} = (1 - \rho(x_i^{n_{co}}, x_o^{n_{co}})) \tag{4}$$

takes all N_{co} pairs $\mathbf{p}$ of images x_i, x_o, $i \neq o$ in the image set $\mathbf{x}$ and forces the corresponding pairs of feature vectors $f_{aux}(x_i), f_{aux}(x_o)$ to have a Euclidean distance $\Delta_{i,o,\mathbf{w}_v}^{n_{co}}$ of exactly $M_{i,o}^{n_{co}}$. In equation 4, $M_{i,o}^{n_{co}}$ is a colour margin that is small for highly correlated colour distributions, leading to small distances $\Delta_{i,o,\mathbf{w}_v}^{n_{co}}$, whereas a low degree of colour correlation results in a large margin, leading to large distances $\Delta_{i,o,\mathbf{w}_v}^{n_{co}}$. Thus, by assuming images of the same class to be semantically and visually similar, integrating the colour similarity loss as well as the semantic similarity loss into network training is supposed to lead to feature clusters that reflect intra-class connectivity and inter-class separability.

3.3 Minibatch Generation

Different kinds of mini-batches for training, i.e. of the image sets $\mathbf{x}$ mentioned above, are compared; randomly drawn mini-batches and class-balanced mini-batches. In the first case, all samples consisting of an image and an assigned class label are randomly drawn from the training dataset, leading to mini-batches with a class distribution similar to the one of the entire dataset. In the latter case, a class label is uniformly drawn first, and then an image is drawn randomly from all training images belonging to the selected class. Thus, the class distribution is approximately uniform in class-balanced mini-batches. In both cases, drawing is performed without replacement to obtain a larger variability of samples in every batch, especially for underrepresented classes.

As a consequence of drawing without replacement, there is a constraint for the mini-batch size N_{MB}. It has to be selected such that the least frequent class c_{min} in the training data will contribute to the loss calculation approximately as often as more frequent classes. Having in total N_{min} examples of c_{min} in the training data and K classes to be distinguished, the batch size is restricted to

$$N_{MB} \leq N_{min} \cdot K. \tag{5}$$

A larger batch size would result in an imbalanced mini-batch even in the case of the class-balanced sampling strategy, because in imbalanced datasets, N_{min} can be smaller than N_{MB}/K for a mini-batch size that does not fulfill the constraint in equation 5.

4. DATASETS

In the context of the EU H2020 project SILKNOW, a knowledge graph consisting of silk records was built and published (Schleider et al., 2021). Each record describes a silk object by means of semantic annotations and potentially one or more images depicting the silk artifact, where only plain fabrics are considered in the present work. Annotations are available for different semantic variables, namely *material*, *time*, *technique* and *place*. These annotations are interpreted as class labels and the number of classes varies between the semantic variables. This is also true for the number of images with a class label for a specific semantic variable. The class structures and the class distributions of the whole dataset are presented in Table 1. The table shows that all of the class distributions are very imbalanced. The number of classes K varies between three for the variable *material* and 29 for *place*.

The empirical class distributions of the individual variables vary with respect to the *imbalance ratio* (IR) and the *imbalance degree*, both of them describing class imbalance (Ortigosa-Hernández et al., 2017). The imbalance-ratio

$$IR = \frac{max_i \zeta_i}{min_j \zeta_j} \tag{6}$$

describes the ratio of the relative frequency ζ_i of examples in the dataset of the most frequent class i and the relative frequency of examples of the most underrepresented class j. Whereas the imbalance-ratio is a suitable measure for describing the imbalance of class distributions for binary classification problems, it does not reflect all characteristics of class distributions for multi-class classification problems as it considers only the frequencies of exactly two classes, i.e. the most frequent and the least frequent ones. In contrast, the imbalance degree also considers the frequencies of other classes in the distribution. We introduce a measure denoted as *balance deviation* (BD) that relies on the imbalance degree proposed in (Ortigosa-Hernández et al., 2017):

$$BD = \frac{d_\Delta(\zeta, \mathbf{e})}{d_\Delta(\zeta, \tau)} \tag{7}$$

In equation 7, $d_\Delta(\cdot)$ is a distance function describing the similarity of two class distributions. We use the total variation

Variable	Class	NS	Class	NS
material	animal fibre	27,252		
	metal thread	4,208		
	vegetal fibre	3,891		
technique	embroidery	6,861	tabby	185
	velvet	3,051	printed	99
	damask	2,768	twill	67
	other technique	2,526	cannele	65
	resist dyeing	355		
place	GB	7,998	RU	228
	FR	7,379	JM	191
	ES	4,708	CH	146
	IT	4,700	EG	117
	IN	2,353	AZ	115
	CN	1,399	MO	84
	IR	1,294	AT	81
	JP	1,097	PT	73
	BE	648	MA	63
	TR	593	BD	60
	DE	592	CA	52
	GR	479	AU	46
	NL	455	MM	46
	US	357	UZ	42
	PK	352		
timespan	19^{th} c.	9,975	16^{th} c.	1,829
	18^{th} c.	8,423	15^{th} c.	685
	20^{th} c.	4,012	13^{th} c.	43
	17^{th} c.	3,378		

Table 1. Statistics of the distribution of samples for the SILKNOW dataset. *Variable*: name of the variable considered; *Class*: classes differentiated for each variables; *NS*: number of samples for a class.

distance (Gibbs and Su, 2002) as similarity function as recommended in (Ortigosa-Hernández et al., 2017), being half of the sum of absolute differences of the two distributions' frequencies. The numerator in equation 7 measures the similarity of the empirical class distribution ζ of a given dataset and the corresponding balanced class distribution $\mathbf{e} := \{\frac{1}{K}, ..., \frac{1}{K}\}$ with K classes. The denominator in equation 7 serves as normalization and expresses the similarity of ζ and a distribution τ that is obtained by eliminating m minority classes, the latter defined to be the classes c with $\zeta_c < \frac{1}{K}$. Thus, τ only has $K - m$ classes with $\zeta_k > 0$ for $k \in \{m + 1, ..., K\}$, and $\sum_{k=m+1}^{K} \zeta_k = 1$; for the the minority classes the frequency is set to zero in τ, i.e. $\zeta_i = 0$ for $i \in \{1, ..., m\}$. Thus, BD is a value in the range of $[0, 1]$ expressing the deviation of ζ from a balanced class distribution. Table 2 contains the statistics about the class distributions of the four variables considered in this paper.

	material	*technique*	*place*	*time*
IR (eq. 6)	7.00	105.55	190.43	231.98
K	3	9	29	7
m	2	5	22	5
m/K	0.67	0.56	0.76	0.71
BD (eq. 7)	0.66	0.91	0.78	0.51

Table 2. Statistics describing the imbalance of the class distributions in Table 1.

5. EXPERIMENTS

The method for training a classifier with an auxiliary clustering loss presented in section 3 is evaluated on the dataset presented in section 4. Section 5.1 gives details on the general experimental setup and the applied evaluation strategy. Section 5.2 presents the results and a discussion of the proposed method.

5.1 Experimental Setup and Evaluation Strategy

In order to train the network in Figure 1, the network weights $\mathbf{w}_{RN}$ are initialized using pre-trained weights obtained on the ILSVRC-2012-CLS dataset (Russakovsky et al., 2015) and the network weights $\mathbf{w}_{tfc}$ and $\mathbf{w}_{sm}$ are randomly initialized using variance scaling (He et al., 2015). During training, the loss function in equation 1 is minimized using mini-batch stochastic gradient descent with adaptive moments (Kingma and Ba, 2014) with the standard parameters ($\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\hat{\epsilon} = 1 \cdot 10^{-8}$). For that purpose, 60% of the data presented in section 4 is used for training and 20% is each used for validation and testing, where all of the images are resized to the input size of the network, i.e. RGB images with 224 x 224 pixels. The network weights are updated based on the training loss, where a batch size of 180 is used, and the validation loss is used for early stopping; training is proceeded until the validation loss is saturated. The hyper parameters are selected based on the F1 scores on the validation set, where preliminary experiments showed that a network with NL = 2 fully connected layers with $[NN^1, NN^2]$ = [1024, 32] nodes is optimal. Further preliminary experiments confirmed a learning rate of $1 \cdot 10^{-3}$ to be optimal.

An overview over all experiments is given in Table 3. Each of the experiments is conducted three times to get a realistic impression of the results despite the random components in training. The experiments in Table 3 are conducted separately for all of the classification tasks, i.e. separate classifiers are trained for the variables *material*, *timespan*, *technique* and *place*. The goal of the experiments is to get an impression of the impact of the individual loss terms on the network's performance to predict the correct class label. A batch size of 180 is selected for all experiments based on the least frequent class in terms of relative frequency among all classification tasks, i.e. class *13^{th} century* of variable *timespan*, and the constraint in equation 5.

Experiment	mini-batch	λ_C	λ_{aux}	α_{sem}	α_{co}
C* (baseline)	rand	1	0	0.0	0.0
C	balanced	1	0	0.0	0.0
C* + sem	rand	1	1	1.0	0.0
C* + co	rand	1	1	0.0	1.0
C* + co + sem	rand	1	1	0.5	0.5
C + sem	balanced	1	1	1.0	0.0
C + co	balanced	1	1	0.0	1.0
C + co + sem	balanced	1	1	0.5	0.5

Table 3. Overview of the experiments conducted for every classification task. *Experiment*: name of the respective experiment. *sampling*: indicates the applied mini-batch sampling strategy described in section 3.3. λ_C, λ_{aux} and α_{sem}, α_{co} refer to the parameters of the loss functions in equations 1 and 2, respectively.

All experiments are evaluated with respect to the *overall accuracy* (OA) and the *F1-scores* achieved on the test set. The OA is the number of the correct predictions in relation to the total number of evaluation examples and thus indicates the percentage of correct predictions. As the OA does not differentiate between individual classes, it can be biased towards the

performance of classes with many examples and does not properly reflect the performance of a classifier in case of imbalanced class distributions. Thus, we also consider the class-wise F1-scores, i.e. the harmonic means of precision and recall, where the precision indicates the percentage of predictions of a class that actually belong to that class and recall indicates the percentage of samples per class in the reference that were correctly assigned to that class by the CNN. The average F1-score of a classification task denotes the mean of all class-specific F1-scores for a variable.

5.2 Results and Discussion

The average quality metrics of all the experiments listed in Table 3 and the corresponding standard deviations are presented in Table 4. The average F1-scores of the m minority classes per variable (cf. section 4) are shown in Table 5. We start with a discussion of some general observations based on the average quality metrics per experiment (section 5.2.1). This is followed by an analysis of the classification performance of the minority classes (section 5.2.2). Finally, the main findings are summarized in section 5.2.3.

5.2.1 Analysis of the average results per variable: Table 4 shows that the overall accuracies and the average F1-scores highly depend on the classification task, i.e. on the variable to be predicted. *Material* obtains both the highest OAs and the highest F1-scores and *place* has the lowest values for both quality metrics. Whereas *technique* has a higher OA compared to *time*, higher F1-scores can be obtained for *time*. Considering the class distributions of the four semantic variables (see Table 2), there seems to be dependency of the OA on the number of minority classes: the smaller the number m of underrepresented classes, the higher the OA. Furthermore, the magnitude of the OA seems to depend on IR: the higher the IR, the lower the OA. The only exception to the latter observation is *place*, which achieves the lowest OA even though its IR is not the highest. This may be due to the high number K of classes to be differentiated for *place*. The number of classes K in a classification task also seems to affect the average F1-scores; *material* with three classes has the highest mean F1-scores ($< 48\%$), followed by *time* with seven classes and scores of up to about 43% and *technique* with nine classes and F1-scores $< 42\%$ and finally, *place* with 29 classes and F1-scores of up to about 21%. Similarly, the magnitude of the largest F1-scores per variable is larger for variables with a smaller number m of underrepresented classes.

Analysing the impact of mini-batch generation on the quality measures in Table 4, the overall accuracies are lower for classifiers trained using mini-batches generated by class-balanced sampling (C) compared to classifiers trained using mini-batches generated by completely random sampling (C^*). The largest decrease in OA of 15.4% can be observed for *place*, followed by *material* with 14.9% and *time*; for the others, the decrease is 3.0% or smaller. The negative impact of the balanced drawing strategy on the OA can be explained by the stronger focus on less frequent classes during training. Consequently, the dominant classes (i.e. the classes having the highest frequencies) have a reduced impact on the determination of the CNN weights, so that they will be more frequently mis-classified. As they have a high impact on the determination of OA, this quality metric will be decreased. Indeed, the magnitude of this decrease seems to depend on the percentage of underrepresented classes in a classification task, m/K (see Table 2); the higher m/K, the larger the difference in OA between the experiments C^* and C. The variable *time* is an exception to this general observation, perhaps because its class distribution is the most balanced one, as indicated by the fact that it has the lowest value of BD.

In contrast to its impact on the OAs, the balanced sampling strategy for generating the mini-batches leads to higher average F1-scores in general, with the exception of *time*, where there is hardly any difference. For *material*, *technique* and *place*, the improvement in the F1 score seems to increase with decreasing values of IR and a decreasing number m of minority classes. A possible reason for the different behaviour of the average F1-scores of *time* could be its comparatively balanced class distribution, indicated by a low value of BD.

Analysing the impact of the auxiliary losses on the quality metrics in Table 4, one can see that the clustering loss focusing on semantic similarity (C^* + *sem*) slightly improves the OAs of *technique* and *time*. All variants of the auxiliary loss lead to a decrease in OA of up to about 1% (for *material* and *place*), which is much less compared to the decrease in OA of up to 15% caused by the class-balanced sampling strategy for gernerating mini-batches. In contrast to the decreases in OA, the F1-scores of all variables are at least slightly improved by at least one variant of the clustering loss. Whereas *material* benefits both from learning semantic similarity and colour similarity (C^* + *co* + *sem*), *technique* only benefits from learning colour similarity (C^* + *co*). This may be the case because *material* and *technique* are more closely related to visual similarity of fabrics than the other variables. *Place* and *time* benefit most from learning semantic similarity (C^* + *sem*), probably because they are more abstract properties; similar colours can probably be found in various places and in different epochs. We assume that especially the underrepresented classes benefit from learning with an auxiliary clustering loss, leading to the improved average F1-scores.

5.2.2 Analysis of the results averaged over the underrepresented classes per variable: A closer look at the performance for the minority classes on the basis of the average F1-scores achieved for these classes, shown in Table 5, basically confirms this assumption. The magnitude of the improvement for the three variables *material*, *place* and *technique* increases with the imbalance of the respective class distribution according to BD. This is true for the improvements caused by balancing the mini-batches (comparison of C^* and C), for the improvements caused by using the auxiliary clustering loss (comparison of C^* and C^* + *aux*, where *aux* can represent any of the losses sem, co, or $sem + co$) and for the improvement of the combined approach (comparison of C^* and C + *aux*). Again, the variable *time* represents an exception, having both the most balanced class distribution, i.e. the lowest value of BD, while at the same time having the largest IR, because it has a single class that is very underrepresented (13^{th} c.). Furthermore, it can be observed that the magnitude of the improvements in the F1-scores of the minority classes of up to about 21% is larger than the improvements in the average F1-scores in Table 4, which are up to 10.5%. This is a strong indication that, as expected, the training modifications primarily support the classifier in correctly predicting underrepresented classes.

5.2.3 Summary: Our results show that the proposed balanced mini-batch generation strategy and the clustering loss improve the classification performance so that the classes are better distinguished by the trained classifier. The class-balanced training strategy results in larger improvements of up to almost 10% in the average F1 score, while the auxiliary clustering

	Overall accuracies [%]				average F1-scores [%]			
Experiment	*material*	*technique*	*place*	*time*	*material*	*technique*	*place*	*time*
C* (baseline)	**76.5** ± 0.25	75.6 ± 0.50	**49.7** ± 0.72	60.6 ± 0.39	37.6 ± 0.37	39.8 ± 1.83	17.6 ± 0.83	41.8 ± 1.75
C	61.6 ± 0.86	73.3 ± 1.89	34.3 ± 2.42	57.6 ± 0.90	47.3 ± 0.17	**41.6** ± 0.92	18.8 ± 0.58	41.4 ± 0.29
C* + sem	76.2 ± 0.48	**75.8** ± 0.51	49.0 ± 0.72	**61.4** ± 0.35	39.2 ± 2.79	39.3 ± 1.09	18.2 ± 1.22	**42.8** ± 0.76
C* + co	75.5 ± 0.60	75.4 ± 0.37	49.1 ± 0.28	60.8 ± 0.99	42.2 ± 1.46	40.0 ± 0.68	18.9 ± 0.36	42.4 ± 1.16
C* + co + sem	75.6 ± 0.26	75.4 ± 0.18	48.5 ± 0.55	**61.4** ± 1.37	42.3 ± 0.75	38.9 ± 1.39	18.6 ± 1.10	42.1 ± 0.95
C + sem	63.1 ± 1.99	74.4 ± 0.85	37.5 ± 1.59	58.2 ± 1.11	47.8 ± 1.43	41.2 ± 0.79	**20.9** ± 0.86	**42.8** ± 0.45
C + co	64.5 ± 0.35	73.5 ± 0.63	38.0 ± 1.51	57.5 ± 1.03	**48.1** ± 0.69	41.5 ± 0.73	20.3 ± 0.70	41.1 ± 0.86
C + co + sem	64.3 ± 1.97	74.1 ± 0.64	37.7 ± 0.76	58.3 ± 0.02	48.0 ± 0.88	41.0 ± 0.04	20.1 ± 0.54	42.3 ± 0.28

Table 4. Overall accuracies [%] and average F1-scores [%] of the experiments. The quality metrics averaged over three runs are presented as well as the corresponding standard deviations. The best average quality metric per variable is highlighted in bold font.

Experiment	*material*	*technique*	*place*	*time*
C* (baseline)	13.1	14.6	8.3	32.2
C	33.8	**19.0**	13.2	32.8
C* + sem	15.6	13.7	9.1	33.4
C* + co	20.3	15.1	10.1	33.1
C* + co + sem	20.5	13.3	9.8	32.4
C + sem	**33.9**	17.5	**15.0**	**34.4**
C + co	33.8	18.5	14.1	32.2
C + co + sem	33.5	17.4	13.9	33.7

Table 5. Average F1-scores [%] over the m minority classes per variable, achieved in three runs.

loss leads to improvements of up to about 5% in that metric. Combining sampling and clustering, the F1-scores can be improved by up to 10.5% on average, where the major improvements can be achieved for underrepresented classes; an analysis of the classification performance of the minority classes shows improvements of up to about 21%. In most of the analysed cases, the improvements in the F1-scores can be attributed to the degree of similarity of the training class distribution to a uniform class distribution, indicated by BD; the exception from this finding ($time$) seems to be related to a high value of IR. However, the better differentiation of the underrepresented classes comes at the cost of a decrease in the performance for the dominant classes and, consequently, a decrease in OA. In case the requirement of the application is to improve the F1 scores without decreasing the OAs, it is recommended to consider the clustering loss during training using randomly sampled mini-batches. In case the focus of the application is on achieving high F1-scores, especially for underrepresented classes, the proposed clustering loss should be applied together with the proposed class-balanced sampling strategy.

6. CONCLUSION AND OUTLOOK

We have presented a training approach for a CNN-based image classifier that combines a classification loss with an auxiliary clustering loss. The clustering loss was supposed to support the classifier to better distinguish the classes that are underrepresented in the training dataset from the others. Furthermore, different strategies for creating mini-batches for training were investigated. The approach was evaluated on four different silk-related classification tasks, where the class distributions of the four tasks varied with regard to the total number of classes, the number of classes with few training examples as well as the degree of the class imbalance. The conducted experiments showed that the proposed training modifications improved the performance of the classifier on all four tasks in terms of the average F1-score. Especially underrepresented classes benefit from both, the balanced sampling strategy and the auxiliary clustering loss. In contrast, the OAs were considerably reduced by applying balanced sampling, whereas training under consideration of the auxiliary loss had hardly any impact on the OAs. Thus, the use of the balanced sampling strategy is only recommended for applications in which a high F1 score for all classes is required while the OA considered to be less relevant.

Future work could focus on other datasets or on further classification approaches. It would be interesting to analyse the behaviour of the presented approach on datasets and tasks that are not related to silk fabrics. Such datasets could contain images depicting other types of artifacts in the domain of cultural heritage, images depicting modern fabrics or clothes or images depicting objects from a completely different domain. Such an analysis would enable a broader analysis of the relation between the classification accuracy and the characteristics of the class distribution. From a methodological point of view, further techniques for addressing the class imbalance problem could be integrated in the approach. This could for example be the utilization of additional synthetic training samples resulting from data augmentation. It is assumed that the SMOTE approach (Chawla et al., 2002) or one of its variants, e.g. (Han et al., 2005; Maciejewski and Stefanowski, 2011), is suitable to mitigate the effects of class imbalance, because they primarily aim to generate synthetic samples for underrepresented classes. Alternatively, introducing an additional augmentation-based similarity loss could help to learn more representative features for images of underrepresented classes. Examples are the self-similarity loss of Dorozynski and Rottensteiner (2022) or the representation learning loss in (Chen and He, 2021). It would be especially interesting to observe the classifier's behaviour if such an augmentation-based similarity loss were primarily applied to samples of underrepresented classes. Finally, the proposed approach should be compared to other existing approaches dealing with class imbalance.

ACKNOWLEDGEMENTS

The research leading to these results is in the context of the "SILKNOW. Silk heritage in the Knowledge Society: from punched cards to big data, deep learning and visual/tangible simulations" project, which has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 769504.

References

Bianco, S., Mazzini, D., Napoletano, P., Schettini, R., 2019. Multitask painting categorization by deep multibranch neural network. *Expert Systems with Applications*, 135, 90–101.

Bishop, C. M., 2006. *Pattern Recognition and Machine Learning*. 1^{st} edn, Springer, New York (NY), USA.

Blessing, A., Wen, K., 2010. Using machine learning for identification of art paintings. Technical Report CS 229, Stanford University, USA.

Caron, M., Bojanowski, P., Joulin, A., Douze, M., 2018. Deep clustering for unsupervised learning of visual features. *Proceedings of the European Conference on Computer Vision (ECCV)*, 132–149.

Caruana, R. A., 1993. Multitask learning: A knowledge-based source of inductive bias. *International Conference on Machine Learning (ICML)*, 41–48.

Chawla, N. V., Bowyer, K. W., Hall, L. O., Kegelmeyer, W. P., 2002. SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16(1), 321–357.

Chen, X., He, K., 2021. Exploring simple siamese representation learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 15750–15758.

Choi, H., Som, A., Turaga, P., 2020. Amc-loss: Angular margin contrastive loss for improved explainability in image classification. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 3659–3666.

Dong, Q., Gong, S., Zhu, X., 2018. Imbalanced deep learning by minority class incremental rectification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(6), 1367–1381.

Dorozynski, M., Clermont, D., Rottensteiner, F., 2019. Multitask deep learning with incomplete training samples for the image-based prediction of variables describing silk fabrics. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, IV-2/W6, 47–54.

Dorozynski, M., Rottensteiner, F., 2022. Deep descriptor learning with auxiliary classification loss for retrieving images of silk fabrics in the context of preserving European silk heritage. *ISPRS International Journal of Geo-Information (IJGI)*, 11(2), 82.

Garcia, N., Renoust, B., Nakashima, Y., 2020. ContextNet: representation and exploration for painting classification and retrieval in context. *International Journal of Multimedia Information Retrieval*, 9(1), 17–30.

Gibbs, A. L., Su, F. E., 2002. On choosing and bounding probability metrics. *International statistical review*, 70(3), 419–435.

Hameed, K., Chai, D., Rassau, A., 2021. Class distribution-aware adaptive margins and cluster embedding for classification of fruit and vegetables at supermarket self-checkouts. *Neurocomputing*, 461, 292–309.

Han, H., Wang, W.-Y., Mao, B.-H., 2005. Borderline-smote: a new over-sampling method in imbalanced data sets learning. *Advances in Intelligent Computing*, 3644, Springer, Berlin, Heidelberg, 878–887.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 1026–1034.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Identity mappings in deep residual networks. *European Conference on Computer Vision (ECCV)*, 630–645.

Huang, C., Li, Y., Loy, C. C., Tang, X., 2016. Learning deep representation for imbalanced classification. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5375–5384.

Johnson, J. M., Khoshgoftaar, T. M., 2019. Survey on deep learning with class imbalance. *Journal of Big Data*, 6(1), 1–54.

Kingma, D. P., Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Krawczyk, B., 2016. Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence*, 5(4), 221–232.

Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012. ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems 25 (NIPS'12)*, 1, 1097–1105.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., Jackel, L. D., 1989. Backpropagation applied to handwritten ZIP code recognition. *Neural computation*, 1(4), 541–551.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. Focal loss for dense object detection. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2980–2988.

Ling, C. X., Sheng, V. S., 2008. Cost-sensitive learning and the class imbalance problem. *Encyclopedia of Machine Learning*, 2011, 231–235.

Liu, X., Vijaya Kumar, B. V. K., You, J., Jia, P., 2017. Adaptive deep metric learning for identity-aware facial expression recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 20–29.

Ma, H., Zhang, Z., Li, W., Lu, S., 2021. Unsupervised Human Activity Representation Learning with Multi-Task Deep Clustering. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(1).

Maciejewski, T., Stefanowski, J., 2011. Local neighbourhood extension of smote for mining imbalanced data. *2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, IEEE, 104–111.

Mani, I., Zhang, I., 2003. Knn approach to unbalanced data distributions: a case study involving information extraction. *Proceedings of the International Conference on Machine Learning (ICML) Workshop on Learning from Imbalanced Datasets*, 126, ICML United States, 1–7.

Nair, V., Hinton, G. E., 2010. Rectified linear units improve restricted boltzmann machines. *Proceedings of the 27th international conference on machine learning (ICML-10)*, 807–814.

Ortigosa-Hernández, J., Inza, I., Lozano, J. A., 2017. Measuring the class-imbalance extent of multi-class problems. *Pattern Recognition Letters*, 98, 32–38.

Pouyanfar, S., Tao, Y., Mohan, A., Tian, H., Kaseb, A. S., Gauen, K., Dailey, R., Aghajanzadeh, S., Lu, Y.-H., Chen, S.-C. et al., 2018. Dynamic sampling in convolutional neural networks for imbalanced data classification. *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, IEEE, 112–117.

Qi, C., Su, F., 2017. Contrastive-center loss for deep neural networks. *2017 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2851–2855.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., Fei-Fei, L., 2015. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision (IJCV)*, 115(3), 211–252.

Schleider, T., Ehrhart, T., Lisena, P., Troncy, R., 2021. SILKNOW knowledge graph. Zenodo. `https://doi.org/10.5281/zenodo.5743090`. Accessed: 2021-11-29.

Sridhar, S., Kalaivani, A., 2021. A survey on methodologies for handling imbalance problem in multiclass classification. *Advances in Smart System Technologies*, 1163, Springer, Singapore, 775–790.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1), 1929–1958.

Strezoski, G., Worring, M., 2017. Omniart: multi-task deep learning for artistic data analysis. *arXiv preprint arXiv:1708.00684*.

Sur, D., Blaine, E., 2017. Cross-depiction transfer learning for art classification. Technical Report CS 231A and CS 231N, Stanford University, USA.

Tan, W. R., Chan, C. S., Aguirre, H. E., Tanaka, K., 2016. Ceci n'est pas une pipe: A deep convolutional network for fine-art paintings classification. *IEEE international conference on image processing (ICIP)*, 3703–3707.

Wen, Y., Zhang, K., Li, Z., Qiao, Y., 2016. A discriminative feature learning approach for deep face recognition. *European Conference on Computer Vision (ECCV)*, 499–515.

Yang, J., Wang, H., Feng, L., Yan, X., Zheng, H., Zhang, W., Liu, Z., 2021. Semantically coherent out-of-distribution detection. *International Conference on Computer Vision (ICCV)*, 8301–8309.

Yang, Z., Liu, T., Liu, J., Wang, L., Zhao, S., 2020. A novel soft margin loss function for deep discriminative embedding learning. *IEEE Access*, 8, 202785–202794.