

Clasificación de Variedades de Semillas de Trigo usando Visión por Computadora

Javier A. Redolfi^{1,2,3}, Diego González Dondo¹, Julián A. Pucheta² y Luis R. Canali¹

¹ Centro de Investigación en Informática para la Ingeniería, Universidad Tecnológica Nacional, Facultad Regional Córdoba, Maestro López S/N, Córdoba, Argentina,

² Universidad Nacional de Córdoba, Haya de la Torre S/N, Córdoba, Argentina,

³ CONICET, Haya de la Torre S/N, Córdoba, Argentina

jredolfi@frc.utn.edu.ar

Resumen En este trabajo abordamos el problema de identificación de variedades de semillas de trigo. La identificación de semillas de trigo es una tarea realizada por personal calificado en diversas etapas de la producción agropecuaria, pero es una actividad lenta, tediosa y de baja repetibilidad. La disponibilidad de un método de clasificación automático de semillas acelera los procesos de evaluación y permite que sean realizados en diferentes etapas del proceso de producción de manera simple y con bajo costo. La solución propuesta es el uso de técnicas actuales de clasificación de imágenes como son Vectores de Fisher de la Familia Exponencial y Redes Neuronales Convolucionales. Con estas técnicas se logra una exactitud del 95 % en la clasificación de un dataset de semillas de 6 variedades de trigo recolectado para esta tarea el cual se encuentra disponible al público para futuras evaluaciones.

Keywords: Agricultura de Precisión, Visión por Computadora, Clasificación de Semillas, Vectores de Fisher, Redes Neuronales Convolucionales

1. Introducción

El análisis de semillas es una actividad muy importante en el proceso de producción de granos. Esta actividad debe ser realizada en diferentes etapas del proceso, incluyendo la producción de semillas, la calificación del cereal para industrialización o comercialización y también durante las investigaciones científicas para mejoras de las especies [6]. En la Argentina por reglamentación, antes de la comercialización de distintos tipos de granos se debe realizar un análisis de una pequeña muestra del lote a comercializar [7].

La tarea de identificación es realizada por personal capacitado usando inspección visual [13], pero en la mayoría de los casos los métodos usados son lentos, tediosos [15], tienen baja reproducibilidad y agregan un grado de subjetividad difícil de cuantificar [6]. Esto se debe entre otras cosas a que el perito debe separar físicamente los granos e identificar el tipo de semilla [13].

Esta actividad es clave para contribuir al agregado de valor al cultivo [6] y también porque el uso final de los granos depende del tipo y variedad de semilla específico [16]. En la actualidad el uso de semillas certificadas juega un rol muy importante en el incremento de la calidad y cantidad del cultivo [17], por lo tanto antes de sembrar una variedad de semillas es muy importante confirmar la variedad a utilizar [17]. El análisis también nos puede brindar conocimiento adicional sobre el proceso de producción, control de la calidad de las semillas y en la identificación de las impurezas [15].

Por lo dicho es de gran importancia técnica y económica la implementación de métodos automáticos basados en visión por computadora para una clasificación confiable y rápida de semillas [6]. Estos métodos también pueden ser explotados para detectar semillas infectadas por insectos o para detectar granos dañados [2].

En este trabajo se propone el uso de visión por computadora para clasificar semillas de distintas variedades de trigo.

Las contribuciones de esta investigación son las siguientes:

- Un sistema automático capaz de clasificar semillas de distintas variedades de una misma especie, en especial trigo,
- basado en técnicas actuales de clasificación de imágenes
- y un dataset de imágenes de semillas de trigo con 6 variedades distintas.

2. Problema

Nos proponemos abordar el problema de la identificación de la variedad de una semilla de trigo dada una imagen de la misma. Esta tarea es muy importante en diversas etapas del proceso de producción y obligatoria en otras. La evaluación es realizada por personal capacitado, pero es lenta, tediosa y de baja reproducibilidad. La posibilidad de automatizar este proceso hace que sea más rápido y preciso, permitiendo una mejor determinación de la calidad de un lote y además que este proceso sea aplicado en varias etapas de la producción, lo cual le da un valor agregado al producto. También la identificación automática le permite a personas no entrenadas validar la procedencia de un lote de semillas en forma simple, posiblemente en una aplicación corriendo sobre un teléfono móvil.

En este trabajo proponemos evaluar el comportamiento de técnicas actuales de visión por computadora como Vectores de Fisher de la Familia Exponencial (eFV) [20] y Redes Neuronales Convolucionales (CNN) [9] para resolver el problema de la identificación de la variedad de una semilla de trigo dada una imagen de la misma.

Estos algoritmos ya han sido usados, dando buenos resultados, en problemas de clasificación de imágenes en general como ImageNet [21] y en problemas de clasificación de subespecies, como por ejemplo aves [26] o plantas [20]. En estos últimos tipos de problemas, conocidos como de grano fino, al igual que en el de identificación de variedades de semillas de la misma especie, las diferencias entre las clases son muy sutiles y escapan al ojo humano no entrenado. En la sección

de resultados demostramos que este tipo de algoritmos se comporta muy bien en este problema en particular.

3. Solución Propuesta

Para resolver el problema de identificación de la variedad de una semilla de trigo planteamos el uso de algoritmos actuales utilizados para clasificación de imágenes. Los algoritmos elegidos son eFV y CNN.

Un eFV es una representación global de una imagen que se obtiene agrupando características locales de una imagen (sección 3.1). Luego esta característica global, la cual es un vector, es usada como entrada a un clasificador lineal.

Las CNN son un tipo de redes neuronales artificiales del tipo *feed-forward* en las cuales los patrones de conectividad entre las neuronas están inspirados en la organización de la corteza visual animal (sección 3.2).

3.1. Vectores de Fisher de la Familia Exponencial

La representación usando eFV [23] nos permite obtener un vector global que describe los descriptores locales de una imagen. Esta representación codifica los descriptores locales de una imagen como un vector gradiente que caracteriza la distribución de estos con respecto a los parámetros de un modelo probabilístico. La formulación original de eFV es conocida como Vectores de Fisher (FV) [22] y está basada en un modelo de distribución de probabilidad Gaussiana, con lo cual a priori solo puede codificar descriptores del tipo reales, pero no binarios o enteros. eFV generaliza esta formulación a un modelo más amplio de distribuciones de probabilidad, llamada familia exponencial, la cual permite un mayor número de dominios de entrada (descriptores) como reales, enteros, binarios o matrices de covarianza.

Supongamos que tenemos una imagen I sobre la que se computan un conjunto de descriptores \mathbf{X}_I en forma densa. Sea $P(\mathbf{X}|\lambda)$ la distribución de probabilidad que modela la distribución de descriptores genéricos computados sobre un conjunto grande de imágenes y $\log P(\mathbf{X}|\lambda)$ la verosimilitud. En donde λ son los parámetros de la distribución.

El eFV de la imagen I es computado como el vector de gradientes con respecto a los parámetros de la función de verosimilitud de los descriptores \mathbf{X}_I extraídos en la imagen

$$g(\mathbf{X}_I) \triangleq L_\lambda \nabla_\lambda \log P(\mathbf{X}_I|\lambda) \quad (1)$$

L_λ es un normalizador obtenido como la descomposición de Cholesky de la inversa de la matriz de información de Fisher.

La clasificación usando eFV contiene varias etapas. En la figura 1 vemos un diagrama completo con las diferentes partes. La entrada es la imagen a la cual le

extraemos descriptores visuales en forma densa (DSCR); luego reducimos estos descriptores en dimensionalidad utilizando Análisis de Componentes Principales (PCA) [1]; a estos nuevos descriptores los codificamos usando eFV y por último clasificamos los vectores usando SVM [4]. Como se ve en la figura, para calcular PCA y eFV necesitamos dos modelos (modelo PCA, modelo eMM); estos modelos los entrenamos en forma no supervisada sobre el conjunto de entrenamiento. Para el cómputo de estas etapas utilizamos la librería presentada en [23].

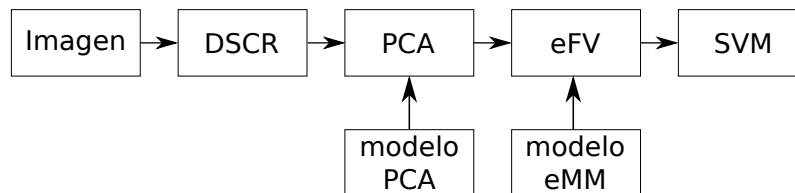


Figura 1: Diagrama en bloques de la codificación usando eFV.

3.2. Redes Neuronales Convolucionales

Las CNN son variantes del perceptrón multicapa inspiradas biológicamente, las cuales tratan de imitar la corteza visual animal. Se sabe por estudios que la corteza visual contiene un arreglo complejo de células sensitivas a pequeñas regiones del campo visual. Estas células se agrupan formando campos receptivos [8] que son emulados en las CNN con filtros de convolución.

Estos tipos de redes son usadas desde hace mucho tiempo, por ejemplo para reconocer imágenes de dígitos manuscritos [11], pero su uso en otras aplicaciones se ha visto relegado en el tiempo debido a que este tipo de modelos son muy lentos para entrenar por el gran número de parámetros a ajustar y por la gran cantidad de imágenes de entrenamiento que se necesitan para lograr un correcto funcionamiento.

Pero relativamente hace pocos años con la aparición de grandes dataset de imágenes etiquetadas como ImageNet [3] y la disponibilidad de Unidades de Procesamiento Gráficas (GPU) muy potentes, estos tipos de modelos se convirtieron en el estándar de facto para aplicaciones de Visión por Computadora, aunque también se usan en otros ámbitos como Procesamiento del Lenguaje Natural, Identificación de Drogas y otras aplicaciones de Machine Learning e Inteligencia Artificial. [10]

Una CNN está compuesta por el apilado de diferentes capas. Las primeras son capas de convolución las cuales aplican filtros de convolución con parámetros entrenables. Entre estas capas además hay otras capas que realizan rectificaciones lineales y agrupamiento (*pooling*) de datos para ir reduciendo la dimensionalidad de la entrada y la complejidad. Luego de estas capas de convolución tenemos

otras capas, como las que normalmente se usan en redes neuronales, conocidas como capas totalmente conectadas. Estas últimas están compuestas por neuronas totalmente conectadas.

Una red neuronal *feed-forward* puede ser vista como el cómputo de una composición de funciones en donde cada función representa una capa, la entrada es la imagen y la salida es un *score* de clasificación para cada una de las clases con la cual fue entrenada. Sean \mathbf{X} los valores de los pixeles de la imagen de entrada, f_l la función que computa la capa l y \mathbf{w}_l el vector de parámetros de dicha función, la salida de la CNN se obtiene como:

$$\mathbf{y} = f(\mathbf{X}) \triangleq f_L(\dots f_2(f_1(\mathbf{X}; \mathbf{w}_1); \mathbf{w}_2); \mathbf{w}_L) \quad (2)$$

El vector de salida \mathbf{y} es un vector que pertenece a \mathbb{R}^c en donde c es el número de clases y luego utilizando la función $\text{argmax}(\mathbf{y})$ computamos cual es la clase que más se ajusta a la imagen de entrada.

Para aprender los parámetros \mathbf{w} de la red, $\mathbf{w} = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_L)$, se debe minimizar una función de riesgo empírico utilizando un conjunto de entrenamiento. La función de riesgo a minimizar es:

$$L(\mathbf{w}) = \frac{1}{n} \sum_{i=1}^n l(z_i, f(\mathbf{X}, \mathbf{w})) \quad (3)$$

en donde $l(z, \hat{z})$ expresa la penalización de predecir la clase \hat{z} en vez de z y n es el número de muestras. Para minimizar L se utiliza el algoritmo de descenso de gradiente:

$$\mathbf{w}^{t+1} = \mathbf{w}^t - \eta_t \frac{df}{d\mathbf{w}}(\mathbf{w}^t) \quad (4)$$

Este algoritmo actualiza el peso actual \mathbf{w}^t en la dirección de mayor descenso de la función de riesgo $L(\mathbf{w})$ computando su gradiente. En dicha ecuación η_t indica la tasa de aprendizaje.

Estos tipos de redes tienen alrededor de 100k parámetros, por lo tanto para su entrenamiento es necesario contar con un dataset de entrenamiento como mínimo 10 veces mayor al número de parámetros. Como normalmente no contamos con esa cantidad de imágenes de entrenamiento sobre el problema a resolver y tampoco resulta factible construirlo, la solución normal es utilizar una red ya entrenada en otro dataset y luego usando el dataset específico al problema, reentrenar solo algunas capas de la red, específicamente las capas totalmente conectadas.

En la figura 2 se ve la configuración de las distintas capas de la CNN utilizada en este trabajo. La entrada a la red es la imagen (IMG) a la que le aplicamos la primer capa de convolución (C1), luego sobre esta, una capa de agrupamiento (P1) y una de rectificación o normalización (N1). Luego vienen otras capas de convolución (C2-C5), de agrupamiento (P2 y P5) y de rectificación (N2). Por

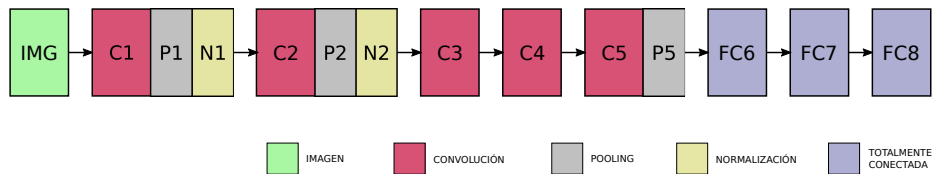


Figura 2: Configuración de la CNN utilizada.

último computamos las capas totalmente conectadas (FC6-FC8). La última de estas capas es una medida de confianza para cada una de las clases.

Esta red es conocida como AlexNet [10] y fue entrenada para la clasificación de un subconjunto de 1 millón de imágenes del dataset ImageNet las cuales pertenecen a 1000 diferentes clases de objetos. Luego con el dataset de semillas de variedades de trigo (sección 4.1) se reentrenaron las últimas capas de la red para ajustarlas al problema. Este reentrenamiento es conocido como ajuste fino o *fine-tuning*. El entrenamiento y la clasificación de la red fue realizada con el framework caffe [9].

3.3. CNN como extractor de descriptores

Otra opción posible de uso de las CNN es para la extracción de descriptores en imágenes [19]. Los descriptores se pueden usar en forma similar a los descriptores comunes usados en la literatura como por ejemplo SIFT pero a diferencia de este los parámetros con los que se calculan no son fijados a mano, sino que son aprendidos en grandes dataset de imágenes. Para la extracción de estos descriptores se utiliza la misma red que se muestra en la figura 2, pero se toma como salida alguna de las capas internas. Lo más común es tomar como descriptor la salida de la segunda capa totalmente conectada, FC7 en este caso. Estos descriptores serán referenciados en lo siguiente como CNND.

Estos descriptores son usados de dos formas en este trabajo. La primera es para computar un *baseline* de clasificación como se propone en [19]. Para este *baseline* se computa un descriptor CNND sobre las imágenes y luego se entrena un SVM para clasificarlos. La segunda forma de usar estos descriptores es codificándolos con el método de eFV que se explicó anteriormente.

4. Experimentos

Para evaluar la factibilidad del esquema propuesto se realizaron experimentos con los diferentes algoritmos presentados en la sección anterior sobre un dataset construido para esta tarea.

4.1. Dataset

Uno de los aportes de este trabajo es la construcción de un dataset de imágenes de semillas de trigo con su correspondiente etiqueta para su uso en la evaluación de algoritmos de clasificación. Dicho dataset se encuentra disponible para descarga en la página web del primer autor ⁴. Hasta nuestro conocimiento este es el primer dataset de este tipo disponible para descarga el cual permite realizar experimentos y poder comparar los resultados de los diferentes algoritmos. El mismo está compuesto por 315 muestras de 6 variedades diferentes de trigo las cuales distinguimos como variedad 1 a 6. Estas muestras (semillas) fueron clasificadas por profesionales en la materia. En la tabla 1 mostramos el número de muestras para cada variedad que contiene el dataset.

Tabla 1: Cantidad de muestras por variedad.

Variedad	1	2	3	4	5	6	Total
# muestras	57	59	52	35	29	83	315

En la imagen 3 vemos ejemplos de las 6 variedades que conforman el dataset. Como se puede ver en la imagen las diferencias entre las variedades de semillas son casi imperceptibles para una persona no entrenada.

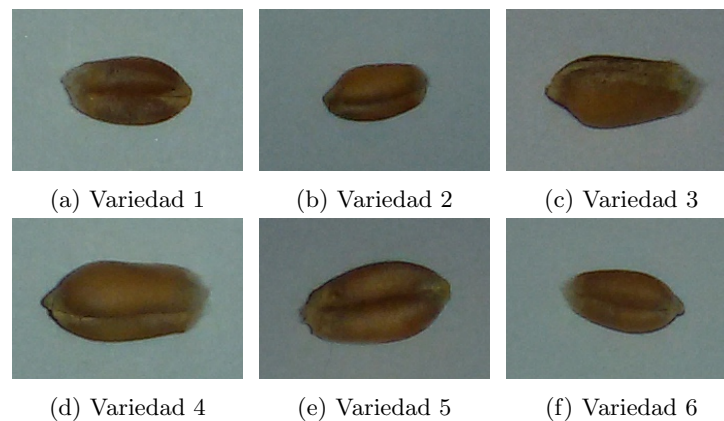


Figura 3: Muestra con las 6 variedades que componen el Dataset

⁴ <http://ciiii.frc.utn.edu.ar/JavierAndresRedolfi>

4.2. Configuraciones

Como ya comentamos en la sección 3.3 el *baseline* de evaluación es la clasificación de descriptores DCCN con SVM, este método se designará como DCNN+SVM.

También evaluamos la clasificación usando eFV presentada en la sección 3.1 con diferentes descriptores. Los descriptores evaluados son SIFT [12], DCOV [24] y DCNN (sección 3.3) los cuales designaremos DSIFT+eFV, DCOV+eFV y DCNN+eFV respectivamente.

Por último como planteamos en la sección 3.2 ajustamos las últimas capas de la CNN AlexNet para el problema en particular; este método se indicará como CNN.

Para el entrenamiento de los diferentes algoritmos se dividió al dataset en 2 partes, una de evaluación y otra de entrenamiento. La parte de entrenamiento está compuesta por 20 imágenes de cada una de las variedades, las cuales hacen un total de 120 imágenes y las restantes 195 imágenes se usaron para evaluación.

4.3. Código

Los cálculos relacionados con eFV se hicieron con la librería **vrl** [23] y para los relacionados con CNN se utilizó la librería **caffe** [9].

4.4. Resultados

Para comparar los diferentes métodos propuestos evaluamos la exactitud de cada uno de ellos en el conjunto de evaluación, el cual está compuesto por 195 muestras. La exactitud la computamos como:

$$Exactitud = 100 \frac{Correctas}{Totales} \quad (5)$$

en donde Correctas es el número de muestras bien clasificadas y Totales es el número de muestras que conforman el dataset.

En la tabla 2 vemos las comparaciones de la exactitud para cada método de clasificación propuesto.

El método basado en CNN fue el que mejor se comportó para este problema superando en 8 puntos al mejor resultado publicado en la literatura [16], en el cual clasifican imágenes de 6 variedades de trigo. Este resultado era el que esperábamos viendo el comportamiento de estas redes en otros problemas de visión por computadora [5, 19] y clasificación de imágenes en particular [10]. Aunque estas redes nunca habían sido utilizadas para este problema en particular.

En la figura 4 mostramos un gráfico de la matriz de confusión normalizada para el método con el cual se obtuvo mayor exactitud (CNN). En el eje vertical tenemos las variedades verdaderas y en el horizontal las predichas. Como vemos la variedad en la que más error cometemos es en la primera.

Tabla 2: Resultados obtenidos con los diferentes métodos.

Método	Exactitud (%)
DCNN+SVM	83.59
DCOV+eFV	91.79
DSIFT+eFV	81.54
DCNN+eFV	92.82
CNN	95.42

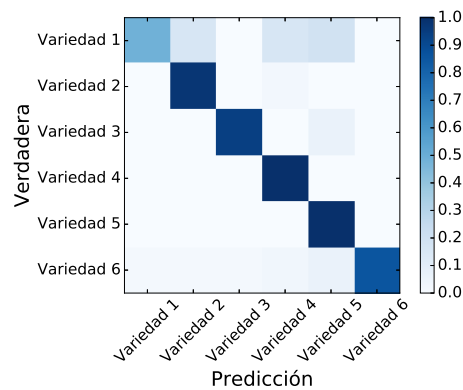


Figura 4: Matriz de confusión para la clasificación usando CNN.

5. Trabajos Relacionados

Existe bastante literatura sobre clasificación de semillas a través de imágenes los cuales podemos agrupar dependiendo del tipo de semillas que clasifican, los descriptores que utilizan y el sensor con el cual capturan las imágenes. La mayoría de ellos utilizan un *pipeline* de clasificación de imágenes general.

En algunos de los trabajos la clasificación se realiza entre semillas de distintas especies [7, 14, 15], entre semillas de una especie y contaminantes que pueden afectar la calidad de la misma [13, 18] y en otros la clasificación es entre variedades de la misma especie [2, 16, 17, 25]. La clasificación de variedades de la misma especie es la que mayor dificultad tiene porque muestras de diferentes clases presentan una gran similitud, en cambio muestras de diferentes especies normalmente tienen una gran diferencia visual.

También podemos encontrar diferentes sensores para adquirir las imágenes a clasificar, en la mayoría se usan cámaras RGB, pero en otros se utilizan cámaras NIR (Infrarojo Cercano) [18, 25]. Con el uso de cámaras del tipo NIR se obtienen

resultados levemente mejores, pero con la desventaja de que éstas son menos comunes y más caras que las RGB.

Dependiendo del tipo de sensor, podemos diferenciar a los trabajos según el tipo de descriptores que utilizan, para cámaras RGB las más usadas son de color, morfológicos y de textura [2,7,13–17]. Para sensores NIR se utilizan descriptores obtenidos con técnicas espectrales [18,25].

Una de las desventajas de los trabajos analizados a diferencia del que presentamos, es que en ninguno de ellos se pone a disposición para descarga el dataset con el cual se realizan los experimentos. La disponibilidad de los dataset con los cuales se realizan los experimentos permiten que la comparación entre los distintos métodos sea más simple y rápida, contribuyendo al avance científico en el área.

Con respecto a la exactitud de los diferentes métodos analizados, los que discriminan entre semillas de diferentes variedades o semillas de sus contaminantes, el cual es un problema más simple tienen una exactitud de entre 95 % a 100 %. En los trabajos en los cuales se discrimina entre variedades de la misma semilla [2,16,17,25], la exactitud es un poco menor y va entre 87 % a 98 %. Pero estos últimos trabajos, que obtienen una exactitud del 98 % no son directamente comparables con los que obtuvimos en este trabajo porque en [25] las imágenes son capturadas con una cámara NIR y en [2] no se clasifican imágenes de una sola semilla de trigo, sino que se clasifican imágenes con muchas semillas de trigo de la misma variedad.

6. Conclusiones y Trabajo a Futuro

En este trabajo se propuso el uso de técnicas actuales de clasificación de imágenes y visión por computadora como son eFV y CNN para abordar el problema de identificación de variedades de semillas de trigo. Con el uso de CNN se logró una exactitud en la identificación del 95 % lo cual demuestra la potencialidad de estos modelos. Además se contribuyó con un dataset de 6 variedades de trigo el cual se encontrará disponible para descarga en la página web del primer autor.

Como trabajo a futuro planeamos evaluar la capacidad del algoritmo para detectar defectos en las semillas como manchas, semillas partidas y también detectar partículas extrañas en las muestras, como podrían ser otras semillas, cáscaras, etc. Otra opción interesante a evaluar es la detección de enfermedades y evaluación de la calidad de las muestras. Para evaluar estas nuevas propuestas planeamos aumentar el dataset con estos tipos de muestras.

7. Agradecimientos

El primer autor agradece al Centro de Investigación en Informática para la Ingeniería en el cual desarrolla sus tareas como becario doctoral del CONICET.

Para el cómputo se utilizaron las computadoras ZX81 y JupiterAce de la FaMAF-UNC, adquiridas con fondos de PROMINF y PAMEG y placas aceladoras donadas por NVIDIA al Grupo de GPGPU Computing de FaMAF-UNC.

Referencias

1. Bishop, C.M.: Pattern recognition and machine learning. springer (2006)
2. Chaugule, A., Mali, S.N.: Evaluation of texture and shape features for classification of four paddy varieties. *Journal of Engineering* 2014 (2014)
3. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. pp. 248–255. IEEE (2009)
4. Fan, R.E., Chang, K.W., Hsieh, C.J., Wang, X.R., Lin, C.J.: Liblinear: A library for large linear classification. *The Journal of Machine Learning Research* 9, 1871–1874 (2008)
5. Girshick, R.: Fast r-cnn. In: *Proceedings of the IEEE International Conference on Computer Vision*. pp. 1440–1448 (2015)
6. Granitto, P.M., Navone, H.D., Verdes, P.F., Ceccatto, H.: Weed seeds identification by machine vision. *Computers and Electronics in Agriculture* 33(2), 91–103 (2002)
7. Granitto, P.M., Verdes, P.F., Ceccatto, H.A.: Large-scale investigation of weed seed identification by machine vision. *Computers and Electronics in Agriculture* 47(1), 15–24 (2005)
8. Hubel, D.H., Wiesel, T.N.: Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology* 195(1), 215–243 (1968)
9. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding. In: *Proceedings of the ACM International Conference on Multimedia*. pp. 675–678. ACM (2014)
10. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. pp. 1097–1105 (2012)
11. Le Cun, B.B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D.: Handwritten digit recognition with a back-propagation network. In: *Advances in neural information processing systems*. Citeseer (1990)
12. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International journal of computer vision* 60(2), 91–110 (2004)
13. Luo, X., Jayas, D., Symons, S.: Identification of damaged kernels in wheat using a colour machine vision system. *Journal of cereal science* 30(1), 49–59 (1999)
14. Majumdar, S., Jayas, D.: Classification of bulk samples of cereal grains using machine vision. *Journal of Agricultural Engineering Research* 73(1), 35–47 (1999)
15. Pandey, N., Krishna, S., Sharma, S., et al.: Automatic seed classification by shape and color features using machine vision technology. *International Journal of Computer Applications Technology and Research* 2(2), 208–213 (2013)
16. Pazoki, A., Pazoki, Z.: Classification system for rain fed wheat grain cultivars using artificial neural network. *African Journal of Biotechnology* 10(41), 8031–8038 (2013)
17. Pourreza, A., Pourreza, H., Abbaspour-Fard, M.H., Sadrnia, H.: Identification of nine iranian wheat seed varieties by textural analysis with image processing. *Computers and electronics in agriculture* 83, 102–108 (2012)

18. Ravikanth, L., Singh, C.B., Jayas, D.S., White, N.D.: Classification of contaminants from wheat using near-infrared hyperspectral imaging. *Biosystems Engineering* 135, 73–86 (2015)
19. Razavian, A.S., Azizpour, H., Sullivan, J., Carlsson, S.: CNN Features off-the-shelf: an Astounding Baseline for Recognition. In: *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2014 IEEE Conference on. pp. 512–519. IEEE (2014)
20. Redolfi, J.A., Sánchez, J.A., Pucheta, J.A.: Fisher vectors for leaf image classification: An experimental evaluation. In: *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, pp. 298–305. Springer (2015)
21. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: Imagenet large scale visual recognition challenge. *International Journal of Computer Vision* 115(3), 211–252 (2015)
22. Sánchez, J., Perronnin, F., Mensink, T., Verbeek, J.J.: Image Classification with the Fisher Vector: Theory and Practice. *International Journal of Computer Vision* 105(3), 222–245 (2013)
23. Sánchez, J., Redolfi, J.: Exponential family Fisher vector for image classification. *Pattern Recognition Letters* 59, 26–32 (2015)
24. Tuzel, O., Porikli, F., Meer, P.: Region covariance: A fast descriptor for detection and classification. In: *Computer Vision–ECCV 2006*, pp. 589–600. Springer (2006)
25. Wang, D., Dowell, F., Lacey, R.: Single wheat kernel color classification by using near-infrared reflectance spectra. *Cereal chemistry* 76(1), 30–33 (1999)
26. Zhang, N., Donahue, J., Girshick, R., Darrell, T.: Part-based r-cnns for fine-grained category detection. In: *Computer Vision–ECCV 2014*, pp. 834–849. Springer (2014)