

RGB Images Driven Recognition of Grapevine Varieties

Pavel Škrabánek¹[0000-0001-6194-0467], Petr Doležel²[0000-0002-7359-0764],
Radomil Matoušek¹[0000-0002-3142-0900], and Petr Junek¹

¹ Brno University of Technology, Brno, Czech Republic,
`pavel.skrabane@vut.cz`

² University of Pardubice, Pardubice, Czech Republic

Abstract. We present a grapevine variety recognition system based on a densely connected convolutional network. The proposed solution is aimed as a data processing part of an affordable sensor for selective harvesters. The system classifies size normalized RGB images according to varieties of grapes captured in the images. We train and evaluate the system on in-field images of ripe grapes captured without any artificial lighting, in a direction of sunshine likewise in the opposite direction. A dataset created for this purpose consists of 7200 images classified into 8 categories. The system distinguishes among seven grapevine varieties and background, where four and three varieties have red and green grapes, respectively. Its average per-class classification accuracy is at 98.10 % and 97.47 % for red and green grapes, respectively. The system also well differentiates grapes from background. Its overall average per-class accuracy is over 98 %. The evaluation results show that conventional cameras in combination with the proposed system allow construction of affordable automatic selective harvesters.

Keywords: Recognition of grapevine varieties · Densely connected convolutional network · In-field images · Agriculture mechanization.

1 Introduction

In last years, we can observe unprecedented progress of agriculture mechanization towards its full automation. The rapid development in areas such as computer vision and machine learning likewise affordability of powerful hardware and precise manipulators allowed construction of autonomous robotic systems, e.g. for weed control [18], precise spraying [21, 2] and harvesting. Robots capable to crop greenhouse vegetables, apples, grapes [2], sweet peppers [1] and even strawberries [24] have been presented. One of the directions of their further development is selective harvesting. A good example is harvesting of grapevines according to their varieties. The basic prerequisite for such a selective harvester is correct recognition of grapevine varieties.

Recognition of grapevine varieties can be carried out different ways. A traditional recognition method is ampelometry [4]. As the method is visual, it is

non-destructive. However, it requires involvement of an expert with extensive training, even when using a specialized software [20]. Accuracy of this method is strongly dependent on skills and experience of the expert. More objective is for example DNA analysis [15]; however, this method, as well as other wet chemistry techniques, is destructive, time-consuming, labour-demanding and requires involvement of an expert. Automation of grapevine variety identification enabled the current development of computer vision methods and availability of advance image sensors. Methods processing data provided by a spectrometer [5, 3] or a hyperspectral camera [6] are automatic, non-destructive and fast. Measurements of an interaction of electromagnetic radiation with matter at many different spectral bands allow accurate recognition of grapevine varieties. The main disadvantage of this approach is a high purchase price of a spectrometer or a hyperspectral camera, which is several orders of magnitude higher compared to a conventional camera. Implementation of such sensors into a selective harvester would significantly increase its price.

Traditional methods aimed at recognition of grapevine varieties are limited by human senses. For example, ampelometry uses eyesight by the grapevine variety recognition. Despite that the human perceives only visible light in mostly three bands, experts are capable to recognize tens of varieties. Conventional cameras provide images of comparable attributes. We expect that the images keep information that allows an accurate classification of grapes according to their varieties.

An image-based classification of grapevines according to their varieties is a complex task which requires extraction of many discriminative features. An extensive diversity of an outdoor environment further increases the complexity of the feature extraction. The overall complexity of this task requires employment of a state-of-the-art image categorization system.

The state-of-the-art image categorization systems are based on deep convolutional networks (deep ConvNets) [12]. Deep ConvNets allow creation of self-contained image categorization systems which ensure both feature extraction and classification of object images. Key factors influencing performance of such a system are a learning capacity of a deep ConvNet and the quality of a training set. The capacity of the network is given by its topology. Modern topologies control the capacity by varying width or depth of networks [11]. Enlarging a deep ConvNet capacity through increasing its width is used e.g. in GoogLeNet [23]. The second approach is to increase number of network's layers (the network depth), while retaining the data processing linearity. Topologies, such as Highway Networks [22], Residual Networks [8], Deep Pyramidal Residual Networks [7], Densely Connected Convolutional Networks (DenseNets) [9] and Cross-Layer Neurons Networks [26] can have tens to hundreds of layers.

Factors such as selection of training samples, their correct categorization, proportional representation of samples with respect to their categories, as well as the total number of samples in a training set predetermine the quality of the training set [13]. To increase the number of samples, data augmentation

techniques, such as image translations, horizontal reflections [25], and rotations [17] are used.

Herein, we present a variety recognition system. We based the system on a DenseNet topology. A dense connectivity pattern used in DenseNets alleviates a vanishing-gradient problem and it allows creation of very deep networks with high learning capacity [9]. For a training and evaluation of the system, we form a dataset based on in-field photos captured under various lighting conditions.

2 Materials and Methods

2.1 Data Collection

We captured colour photos (individual images) of common grapevine during a harvest. The data collection was carried out within two days in the morning and in the afternoon in August 2015. We selected various locations in vineyards in Čejkovice, Czech Republic. We used no artificial lighting and we captured the photos in a direction of sunshine likewise in the opposite direction (both days was partly sunny). The resulting collection of photos includes Welschriesling, Saint Laurent, Gewürztraminer, Pinot noir, Riesling Weiss, Pinot gris, and Veltliner Grün varieties (names of varieties according to Vitis International Variety Catalogue [14]).

We used camera bodies CANON EOS 1000D and CANON EOS 1100D with CANON ZOOM lenses EF-S 18-55 mm f/3.5-5.6 II and IS II, respectively. Resolutions of the photos are 1936×1288 pixel (px) and 4272×2848 px, respectively. The photos use RGB colour model with 24 bits bit depth. We placed the cameras perpendicular to vineyard rows (in terms of an axis of a lens), in a distance about 1.4 m from the rows, at an altitude of 1.25 m from the ground. A focal length varied between 18 mm and 24 mm.

2.2 Dataset

For training and evaluation of the variety recognition system, we form a dataset of RGB images of resolution 120×120 px. For each variety, we create 900 images. Grape clusters in the images cover at least 70 % of their surface. Further, we create 900 images capturing a background, i.e. the final dataset consists of 7200 images classified into 8 categories (Fig. 1).

The images in the dataset are cut-outs of grapevine photos acquired within the data collection. For this purpose, we randomly select between 12 and 14 photos (depending on a density of grape clusters in photos) of each variety. In Table 1, we provide information about the number of selected photos with respect to grapevine varieties (first column), camera bodies (first row) and focal lengths (second row).

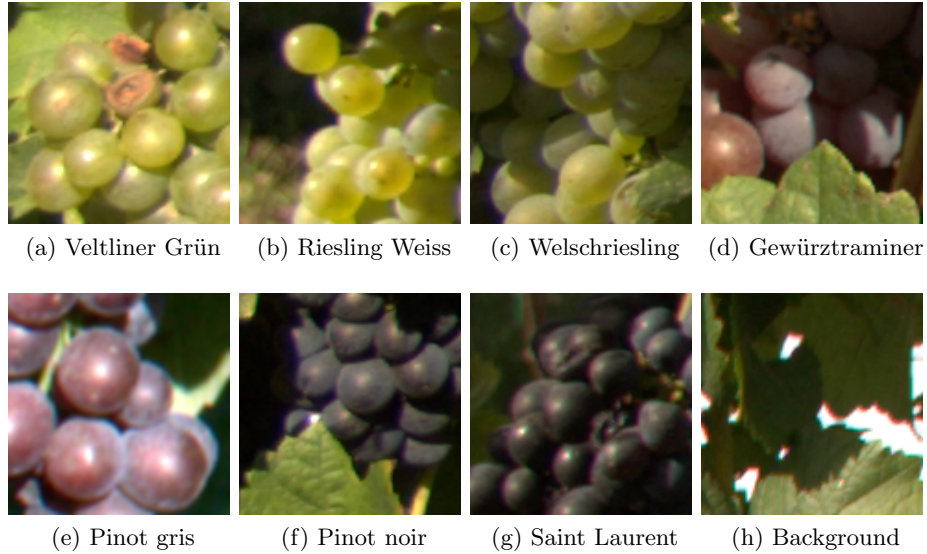


Fig. 1: Example images for the categories in the dataset.

Table 1: **Number of images selected for forming of the dataset.** For each variety (first column), number of used images is stated with respect to the focal length (second row) and the camera body (first row).

| | camera body | | EOS 1000 D | | EOS 1100 D | | |
|----------------|--------------|---|------------|----|------------|----|----|
| | focal length | | 18 | 21 | 24 | 18 | 23 |
| Gewürztraminer | 8 | - | 4 | - | - | - | - |
| Veltliner Grün | 8 | 4 | - | - | - | - | - |
| Pinot gris | 8 | 2 | - | - | - | 4 | - |
| Pinot noir | 8 | - | 4 | - | - | - | - |
| Riesling Weiss | 6 | - | 4 | 4 | - | - | - |
| Saint Laurent | 12 | - | 0 | - | - | - | - |
| Welschriesling | 10 | - | 4 | - | - | - | - |

2.3 Densely Connected Convolutional Networks

As in other deep ConvNets [12], convolutional, pooling and fully connected layers are arranged in a feed-forward manner to form a DenseNet. Regular patterns occurring in DenseNets allow us to simplify description of their topologies. Let us define two composite building elements which will be used to describe a topology of the presented variety recognition system: a *dense blocks* (DB) and a *transition layer* (TL).

Dense Block Let us consider a n -th DB of d_n layers that is built in a DenseNet of L layers. The input and the output of the n -th DB are placed at i_n -th and o_n -th levels of the network, respectively, i.e. $d_n = o_n - i_n + 1$. Feature maps produced at the ℓ -th level of the network, where $\ell \in [i_n, o_n]$, are given as

$$\mathbf{x}_\ell = H_\ell([\mathbf{x}_{i_n-1}, \dots, \mathbf{x}_{\ell-1}]), \quad (1)$$

where $H_\ell(\cdot)$ is a non-linear transformation performed at the ℓ -th level, \mathbf{x}_{i_n-1} are feature maps at the input of the n -th DB, \mathbf{x}_i for $i \in [i_n, \ell - 1]$ are feature maps produced at preceding levels of the n -th DB, and $[\mathbf{x}_{i_n-1}, \dots, \mathbf{x}_{\ell-1}]$ denotes their concatenation.

Two variants of the non-linear transformation $H(\cdot)$ can be used in DBs: a basic and a bottleneck version [9]. The basic version is a composite function which consists of a batch normalization (BN) [10], a rectified linear unit (ReLU), and a convolution (Conv) [12], respectively. Using a short notation, the basic version of $H(\cdot)$ can be written as BN-ReLU-Conv($h \times w, f, s$), where s is stride of convolutional filters, f is number of the filters, and h and w are their height and width, respectively. The bottleneck version of $H(\cdot)$ is defined as BN-ReLU-Conv($1 \times 1, 4f, 1$)-BN-ReLU-Conv($h \times w, f, s$). If necessary, convolutions are zero-padded to keep the feature-map size fixed. For both versions of the composite function $H(\cdot)$, the parameters h, w, s, f are identical for all layers within a DB. We use abbreviations DBa and DBb for DBs with the basic and the bottleneck version of $H(\cdot)$, respectively.

Transition Layer Let us consider a TL connected at the output of the n -th DB (i.e. the TL is placed at the $(o_n + 1)$ -th level of the network). The $(o_n + 1)$ -th TL produces feature maps

$$\mathbf{x}_{o_n+1} = H_{o_n+1}([\mathbf{x}_{i_n-1}, \mathbf{x}_{i_n}, \dots, \mathbf{x}_{o_n}]), \quad (2)$$

where $[\mathbf{x}_{i_n-1}, \mathbf{x}_{i_n}, \dots, \mathbf{x}_{o_n}]$ denotes the concatenation of all feature maps that appear in the n -th DB. H_{o_n+1} is a composite function BN-ReLU-Conv($1 \times 1, f, 1$)-AP($2 \times 2, 2$), where AP($2 \times 2, 2$) denotes an average pooling with pools 2×2 and stride 2 [9].

Compactness of the network is controlled by the number of the 1×1 convolutional filters f incorporated in TLs. The number of feature maps produced by the $(o_n + 1)$ -th TL is given as $f_{o_n+1} = \lfloor \theta m_n \rfloor$, where θ is a compression factor, $\theta \in [m_n^{-1}, 1]$ and m_n is the number of feature maps produced by the n -th DB.

2.4 Variety Recognition System

The presented variety recognition system is a DenseNet. The network classifies RGB images of dimensions 120×120 px according to varieties of grapes captured in the images. We control number of filters f in DBs by a variable k , where $k = 20$. The network is opened by one DBa which consists of one layer ($d = 1$) with $2k$ convolutional filters ($f = 2k$) with kernels of size 7×7 px ($h = w = 7$), stride by 2 px ($s = 2$). The following layer is a max pooling layer (MPL) with pools 3×3 px ($h = w = 3$) stride by 2 ($s = 2$). The inner parts of the network consist of two DBBs with 6 and 9 layers, respectively. At each layer of a DBb, k filters with kernels of size 3×3 px stride by 1 px ensure the feature extraction. Each DBb in the network is followed by one TL. The network is closed by a global average pooling (GAP) and a classifier, respectively. The classifier consists of one fully connected layer of eight neurons followed by a softmax function. We setup the compression factor θ at 0.5. The topology of the network is summarized in Table 2.

Table 2: **Topology of the variety recognition system.** Building elements which form the system are listed with respect to their placement in the network in the first row of the table (the first block is the leftmost one), where DBa and DBb denote the basic and the bottleneck versions of the dense block; MPL is a max pooling layer; TL is a transition layer, GAP denotes a global average pooling, and C is used for a classifier that consists of one fully connected layer followed by a softmax function. The parameters h and w are a height and a weight of a filter kernel or of a pool; s is stride of the kernel or of the pool; f is the number of filters at one convolution in a dense block; and d is the number of layers in the dense block.

| | DBa | MPL | DBb | TL | DBb | TL | GAP | C |
|-----|------|-----|-----|----|-----|----|-----|---|
| h | 7 | 3 | 3 | - | 3 | - | 5 | - |
| w | 7 | 3 | 3 | - | 3 | - | 5 | - |
| s | 2 | 2 | 1 | - | 1 | - | 5 | - |
| f | $2k$ | - | k | - | k | - | - | - |
| d | 1 | - | 6 | - | 9 | - | - | - |

We use MATLAB R2018b and Deep learning toolbox to train and evaluate the system. We split randomly the dataset into a training and an evaluation set, where the training set consists of 750 samples of each category. The rest of images (150 samples of each category) form the evaluation set. We train the system using ADAM optimizer [16] for 500 epochs with mini batches of 400 samples, minimizing a cross entropy function. We setup a learning rate, and an exponential decay rate for first and second moment estimates at 10^{-3} , 0.95 and 0.999, respectively. We shuffle images in the training set every epoch.

We use data augmentation techniques to bring more variability into the training set. We utilize a function `imageDataAugmenter`, where we use a random rota-

tion (range of a rotation angle: ± 20 degree), a random reflection in the left-right direction, a random horizontal and vertical translation (range of a translation distance: ± 3 px), and a random horizontal and vertical shear (range of a shear angle: ± 20 degree).

3 Results and Discussion

We summarize the evaluation results in a confusion matrix (Table 3), where rows and columns represent instances in actual and predicted classes, respectively. We extend the matrix about average per-class accuracy (last column of Table 3). The average per-class accuracy of the i -th class is given as $acc_i = l^{-1} \sum_{i=1}^l \frac{|TP_i| + |TN_i|}{n}$, where l is number of classes, $|TP_i|$ is number of correctly classified samples of the i -th class, $|TN_i|$ is number of correctly classified samples of complementary classes to the class i , and n is total number of samples in the evaluation set [19]. Calculating arithmetic means of the accuracies for red and green grapes, we get accuracies at 98.10 % and 97.47 %, respectively. The overall average accuracy (arithmetic mean of all classes) is at 98.02%.

Table 3: **Confusion matrix.** Rows and columns represent instances in actual and predicted classes, respectively. Average per-class accuracies of the classes are summarized in the last column. Distinguished classes are Gewürztraminer (GT), Veltliner Grün (VG), Pinot gris (PG), Pinot noir (PN), background (BG), Riesling Weiss (RW), Saint Laurent (SL), and Welschriesling (WR).

| | GT | VG | PG | PN | BG | RW | SL | WR | acc |
|----|-----|-----|-----|-----|-----|-----|-----|-----|--------|
| GT | 148 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0.9950 |
| VG | 0 | 140 | 0 | 0 | 1 | 6 | 0 | 3 | 0.9717 |
| PG | 1 | 0 | 141 | 3 | 0 | 0 | 5 | 0 | 0.9875 |
| PN | 0 | 0 | 0 | 139 | 0 | 0 | 11 | 0 | 0.9742 |
| BG | 2 | 0 | 1 | 0 | 146 | 1 | 0 | 0 | 0.9933 |
| RW | 0 | 8 | 0 | 0 | 0 | 140 | 0 | 2 | 0.9783 |
| SL | 1 | 0 | 3 | 17 | 2 | 0 | 127 | 0 | 0.9675 |
| WR | 0 | 16 | 0 | 0 | 1 | 9 | 0 | 124 | 0.9742 |

A detail analysis of the performance using the confusion matrix (Table 3) shows that the system well distinguishes between grapes and background (4 from 150 images of background miss classified as grapes, and 4 from 1050 images of grapes miss classified as background). Confusions occur only among grapes of the same colour (no green grape was classified as red and vice versa).

The system best recognizes Gewürztraminer (only 2 from 150 images of Gewürztraminer miss classified as Pinot gris, and only 4 images of another class miss classified as Gewürztraminer). The second-best recognized variety is Pinot gris (141 from 150 images of Pinot gris correctly classified, and only 6 images of another class miss classified as Pinot gris). For the varieties Pinot noir, Veltliner

Grün, and Riesling Weiss, we observe similar performance (around 140 from 150 images per class classified correctly, and 20, 24 and 16 images of another class miss classified as Pinot noir, Veltliner Grün, and Riesling Weiss, respectively). We observe a confusion between the varieties Veltliner Grün and Riesling Weiss (6 images of Veltliner Grün miss classified as Riesling Weiss, and 8 images of Riesling Weiss miss classified as Veltliner Grün).

The most difficult variety is Saint Laurent (127 from 150 images of Saint Laurent correctly classified). The system mostly confuses this variety with Pinot noir (11 images of Pinot noir miss classified as Saint Laurent, and 17 images of Saint Laurent miss classified as Pinot noir). Also, Pinot gris is mostly miss classified as Saint Laurent (5 from 9 miss classifications). The second problematic variety is Welschriesling (124 from 150 images of Welschriesling correctly classified). The system has difficulty to distinguish Welschriesling from Veltliner Grün (16 miss classifications) and from Riesling Weiss (9 miss classifications). The positive thing is that only 5 images of another class are miss classified as Welschriesling.

4 Conclusion

We show that in-field colour images of ripe grapes acquired by a conventional camera can be used for classification of grapevines according to their varieties. The presented variety recognition system is capable to distinguish among seven grapevines varieties, where four and three varieties have red and green grapes, respectively. The system also well differentiates grapes from background. Its overall average per-class accuracy is over 98 % on images captured without any artificial lighting, in a direction of sunshine likewise in the opposite direction. Considering all these facts, we conclude that the proposed solution allows construction of affordable automatic selective harvesters.

Acknowledgments

The work was supported from ERDF/ESF "Cooperation in Applied Research between the University of Pardubice and companies, in the Field of Positioning, Detection and Simulation Technology for Transport Systems (PosiTrans)" (No. CZ.02.1.01/0.0/0.0/17_049/0008394).

References

1. Bac, C.W., Hemming, J., van Tuijl, B., Barth, R., Wais, E., van Henten, E.J.: Performance evaluation of a harvesting robot for sweet pepper. *Journal of Field Robotics* **34**(6), 1123–1139 (2017). <https://doi.org/10.1002/rob.21709>, <https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.21709>
2. Bontsema, J., Hemming, J., Pekkeriet, E., Saeys, W., Edan, Y., Shapiro, A., Hočevar, M., Oberti, R., Armada, M., Ulbrich, H., Baur, J., Debilde,

- B., Best, S., Evain, S., Gauchel, W., Hellström, T., Ringdahl, O.: CROPS: clever robots for crops. *Engineering & Technology Reference* **1**(1) (2015). <https://doi.org/10.1049/etr.2015.0015>
3. Fernandes, A., Utkin, A., Eiras-Dias, J., Silvestre, J., Cunha, J., Melo-Pinto, P.: Assessment of grapevine variety discrimination using stem hyperspectral data and adaboost of random weight neural networks. *Applied Soft Computing* **72**, 140 – 155 (2018). <https://doi.org/https://doi.org/10.1016/j.asoc.2018.07.059>
 4. Galet, P.: *A Practical Ampelography: Grapevine Identification*. Comstock Pub. Associates, Ithaca, N.Y, 1 edn. (1979)
 5. Gutiérrez, S., Tardaguila, J., Fernández-Novales, J., Diago, M.: Data mining and nir spectroscopy in viticulture: Applications for plant phenotyping under field conditions. *Sensors (Switzerland)* **16**(2) (2016). <https://doi.org/10.3390/s16020236>
 6. Gutiérrez, S., Fernández-Novales, J., Diago, M.P., Tardaguila, J.: On-the-go hyperspectral imaging under field conditions and machine learning for the classification of grapevine varieties. *Frontiers in Plant Science* **9**, 1102 (2018). <https://doi.org/10.3389/fpls.2018.01102>
 7. Han, D., Kim, J., Kim, J.: Deep pyramidal residual networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6307–6315 (July 2017). <https://doi.org/10.1109/CVPR.2017.668>
 8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778 (June 2016). <https://doi.org/10.1109/CVPR.2016.90>
 9. Huang, G., Liu, Z., v. d. Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2261–2269 (July 2017). <https://doi.org/10.1109/CVPR.2017.243>
 10. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: Proceedings of the 32nd International Conference on Machine Learning, ICML 2015. Proceedings of Machine Learning Research, vol. 37, pp. 448–456. PMLR (2015)
 11. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. *Communications of the ACM* **60**(6), 84 – 90 (2017)
 12. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**, 436–444 (May 2015). <https://doi.org/10.1038/nature14539>
 13. Lemnaru, C., Potolea, R.: Imbalanced classification problems: Systematic study, issues and best practices. In: *Enterprise Information Systems*. pp. 35–50. Springer Berlin Heidelberg, Berlin, Heidelberg (2012)
 14. Maul et al.: *Vitis international variety catalogue* (2020), www.vivc.de
 15. Pelsy, F., Hocquigny, S., Moncada, X., Barbeau, G., Forget, D., Hinrichsen, P., Merdinoglu, D.: An extensive study of the genetic diversity within seven french wine grape variety collections. *Theoretical and Applied Genetics* **120**(6), 1219–1231 (2010). <https://doi.org/10.1007/s00122-009-1250-8>
 16. Ruder, S.: An overview of gradient descent optimization algorithms. *CoRR* **abs/1609.04747** (2016)
 17. Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. *Journal of Big Data* **6**(1), 60 (Jul 2019). <https://doi.org/10.1186/s40537-019-0197-0>, <https://doi.org/10.1186/s40537-019-0197-0>
 18. Slaughter, D., Giles, D., Downey, D.: Autonomous robotic weed control systems: A review. *Computers and Electronics in Agriculture* **61**(1), 63 – 78 (2008). <https://doi.org/10.1016/j.compag.2007.05.008>

19. Sokolova, M., Lapalme, G.: A systematic analysis of performance measures for classification tasks. *Information Processing & Management* **45**(4), 427 – 437 (2009). <https://doi.org/10.1016/j.ipm.2009.03.002>
20. Soldavini, C., Schneider, A., Stefanini, M., Dallaserra, M., Policarpa, M.: Super ampelo, a software for ampelometric and ampelographic descriptions in vitis. *Acta Horticulturae* **827**, 253–258 (2009). <https://doi.org/10.17660/ActaHortic.2009.827.43>
21. de Soto, M.G., Emmi, L., Perez-Ruiz, M., Aguera, J., de Santos, P.G.: Autonomous systems for precise spraying – evaluation of a robotised patch sprayer. *Biosystems Engineering* **146**, 165 – 182 (2016). <https://doi.org/10.1016/j.biosystemseng.2015.12.018>
22. Srivastava, R.K., Greff, K., Schmidhuber, J.: Training very deep networks. In: *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*. pp. 2377–2385. NIPS’15, MIT Press, Cambridge, MA, USA (2015)
23. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 2818–2826 (June 2016). <https://doi.org/10.1109/CVPR.2016.308>
24. Xiong, Y., Peng, C., Grimstad, L., From, P.J., Isler, V.: Development and field evaluation of a strawberry harvesting robot with a cable-driven gripper. *Computers and Electronics in Agriculture* **157**, 392 – 402 (2019). <https://doi.org/https://doi.org/10.1016/j.compag.2019.01.009>, <http://www.sciencedirect.com/science/article/pii/S0168169918312456>
25. Xu, Y., Jia, Z., Ai, Y., Zhang, F., Lai, M., Chang, E.I.: Deep convolutional activation features for large scale brain tumor histopathology image classification and segmentation. In: *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. pp. 947–951 (April 2015). <https://doi.org/10.1109/ICASSP.2015.7178109>
26. Yu, Z., Li, T., Luo, G., Fujita, H., Yu, N., Pan, Y.: Convolutional networks with cross-layer neurons for image recognition. *Information Sciences* **433-434**, 241 – 254 (2018). <https://doi.org/https://doi.org/10.1016/j.ins.2017.12.045>