

Transfer Learning study for horses breeds images datasets using pre-trained ResNet Networks

Enrique de la Cal¹, Enol García González¹, and Jose Ramón Villar¹

University of Oviedo, Computer Science Department,
Oviedo, Spain
{delacal,villarjose}@uniovi.es, enolgargon@gmail.com

Abstract. The straight applications of horse breeds images classification by automatic methods can be arranged in three topics: i) morphological research for breeding[1, 3, 17], iii) identifying archaeological deposits[8, 11, 7], and ii) visual identification for automatic sheepling[16, 6, 15].

Current work bears the extension of a previous transfer learning study for comparing common and different horse breeds images from different datasets. In addition the results of the comparison of the different horse breeds images has been compared with the assessment given by a human expert.

After deploying the system on two datasets of 5079 and 670 files, and six and seven breeds, it can stated that the system agree the human expert assesment in almost the 50% of the cases. In addition, the proposal outperforms clearly the results of the baseline system by Atabay for the first dataset.

Keywords: Image Classification · Resnet · Deep Learning · classification algorithms · horse breeds

1 Introduction and motivation

In [5], preliminary work on the Horse World figures in the light of Data Science was presented. This work concluded two important issues: i) a quite complete analysis of the public datasets on horses available on the web (108) was carried out, stating that there are four popular niches: races, images, herd analysis and health, and ii) as an academic challenge, a horse breeds images classification problem was selected from the public datasets analysed, and new results were obtained outperforming the base-line results[2] for this specific problem.

As this is an academic exercise, the straight applications of breeds images classification by automatic methods can be arranged in three topics: i) morphological research for breeding[1, 3, 17], ii) identifying archeological deposits[8, 11, 7], and ii) visual identification for automatic sheepling[16, 6, 15]. Between the former topic, domesticated herd animals

Current work bears the extension of this problem to a transfer learning study between datasets with common and different breeds. This way, a horse images classification system will be deployed for known and unknown breeds.

This work is structured as follows. The following section includes the methodology carried out to deploy the transfer learning study presented, while the experimentation and the discussion of the results are coped in section 3. Finally, conclusions and future work is included in section 4.

2 The proposal

The experimentation proposal is based on two main parts: a first part in which horse breed classification models are created and evaluated, and a second part in which these classifiers are used to analyze the similarity of horse breeds already known with others that the model does not know.

The first part of the proposal is based on taking a set of Deep Learning architectures for image classification and creating models that are able to classify images of 6 horse breeds correctly. To speed up the training of these models, Transfer Learning techniques will be used.

In the second part of the proposal we will make use of the models from the first part. It should be remembered that these models were trained to be able to classify images of 6 specific horse breeds. In this second phase, images of 6 horse breeds will also be used, but these 6 breeds will be different from the breeds of the first. The objective of this phase is to try to use the models from the first phase to determine how similar two horse breeds are. To cross-check this information, an expert in the horse field will manually analyse the horse breeds included in both phases, and evaluate the degree of morphological similarity between the breeds.

Figure 1 summarizes the main ideas of the proposal and lists the architectures to be used in the experimental phase, as well as the breeds used in each of the phases.

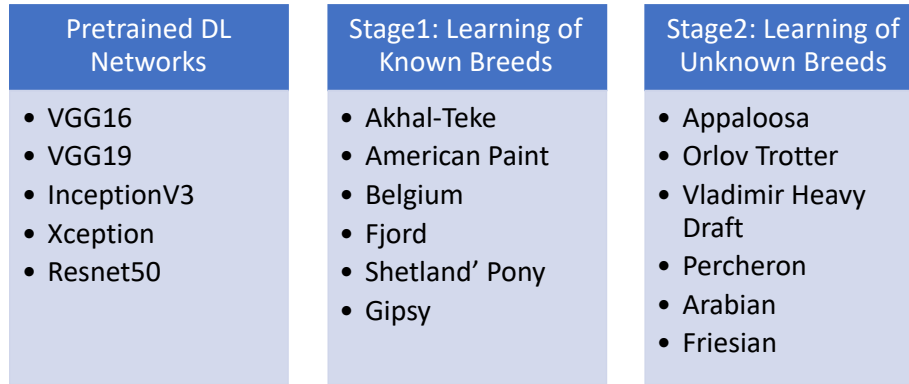


Fig. 1: The general overview of the proposal

3 Numerical results

3.1 Materials and methods

Five pre-trained Deep Learning architectures have been selected: VGG16[13], VGG19[13], Resnet50[9], InceptionV3[14], Xception[4]. Taking these 5 architectures, a transfer learning process has been performed. Three instances of each of these architectures have been initialized with pre-trained network weights taken from ImageNet[12] and a training process has been performed.

For training, the models were set up to be optimized using the stochastic gradient descent method with moment 0.9 and using a learning rate value of $1e^{-1}$. In addition, a Data Augmentation algorithm was used for training in which the images were randomly modified by applying zoom and shear modifications. The shear values vary randomly between 0 and 0.2 degrees and the zoom between 0% and 20%. The number of epochs the training lasts has been set for each architecture manually for each architecture:

- InceptionV3: 100
- Xception: 150
- Resnet50: 100
- VGG16: 200
- VGG19: 200

As, this work proposed using transfer learning, two datasets have been used: i) the source dataset has been taken from Atabay work[2], and the target from Kaggle repository[10]. Let us describe both datasets:

The Atabay dataset, hereafter referred to as D1, consists of 5079 images of 6 horse breeds, containing approximately the same number of images of each breed. To simplify the understanding of the experimentation, numerical values will be used for each breed. The breeds included in this dataset, along with their abbreviations are: Akhal-Teke (D1.B1), American Paint (D1.B2), Belgian (D1.B3), Fjord (D1.B4) Shetland' Pony (D1.B5), Gypsy (D1.B6). Figure 2 includes a sample of the images of each of the breeds contained in D1. This dataset will be the one used in the first part of the proposal to train and perform the test of the models.



Fig. 2: Breeds of Dataset 1

The second dataset used, hereafter referred to as D2, consists of 670 images of 7 horse breeds. As in D1, numerical values will be used for each breed. The breeds included in this dataset, along with their abbreviations are: Akhal-Teke (D2.B1), Appaloosa (D2.B2), Orlov Trotter (D2.B3), Vladimir (D2.B4), Percheron (D2.B5), Arabian (D2.B6), Friesian (D2.B7). Figure 3 includes a sample of the images of each of the breeds contained in D2. This dataset will be the one used in the second part of the proposal, the first breed (Akhal-Teke) is common to D1, which will serve as a control subject.

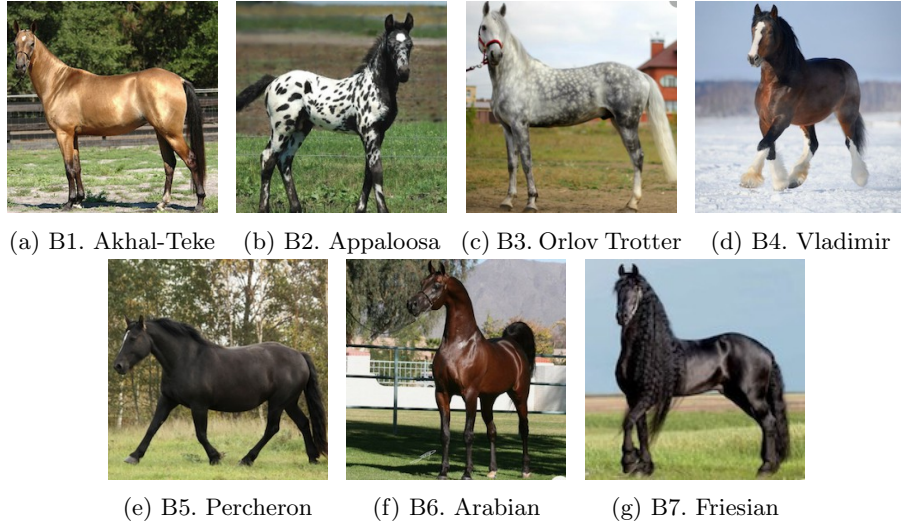


Fig. 3: Breeds of Dataset 2

To have a general idea of the number of images included in each dataset, Table 1 is included, with the breakdown of the number of images in each dataset discriminating according to race, as well as the count and deviation of images between breeds.

Table 1: Number of images contained in each dataset

	D1.B1	D1.B2	D2.B3	D2.B4	D2.B5	D2.B6	Total	Deviation	
<i>D1</i>	876	822	843	810	894	834	5079	32,3341924	
	D2.B1	D2.B2	D2.B3	D2.B4	D2.B5	D2.B6	D2.B7	Total	Deviation
<i>D2</i>	123	105	107	37	56	122	120	670	34,793267

3.2 Results

Thus, the two studies referred in the proposal section are: i) the classification of the known breeds from the datasets D1 as well as the inclusion of the samples D2.B1 in order to check the hardness of the models, and ii) a similarity study between the breeds from D1 and D2. In addition, a discussion section is included comparing the similarity study results obtained for our models for unknown breeds and the corresponding assessment of an human expert.

Comparative performance between our models and Atabay for dataset D1 Table 2 shows the accuracy of our proposal for the six breeds of dataset D1

as well as the whole dataset (All B) compared to Atabay results also for the whole dataset (All B). It can be stated that our proposal outperforms clearly the Atabay results for all the networks architectures (Network) for the whole dataset (All B). Concerning the single breed performance, it can be seen that the worst behaviour is obtained for B3 in all models.

Table 2: Accuracy of our models (Our Proposal columns) compared with Atabay results for the test fold of dataset D1

Network	Our Proposal							Atabay
	B1	B2	B3	B4	B5	B6	All B	All B
InceptionV3	0.9875	0.9714	0.9463	0.9806	0.9947	0.9755	0.9762	0.8879
Xception	0.9893	0.9390	0.9333	0.9690	0.9610	0.9831	0.9626	0.9300
ResNet50	0.9840	0.9829	0.9759	0.9884	0.9699	0.9755	0.9793	0.9590
VGG16	0.9840	0.9771	0.9556	0.9884	0.9734	0.9718	0.9750	0.9069
VGG19	0.9822	0.9657	0.9574	0.9826	0.9876	0.9755	0.9753	0.9005

Besides, the results of table 3 shows that the performance of the proposed models is reduced lightly respect to the original models, after including the images from the D2.B1 dataset into the D1 test dataset, bearing the hardness of the models out.

Table 3: Accuracy for the test fold of dataset D1 including the D2.B1 samples

	D1.B1	D1.B1+D2.B1	B2	B3	B4	B5	B6	All B
InceptionV3	0.9875	0.9882	0.9714	0.9463	0.9806	0.9947	0.9755	0.9775
Xception	0.9893	0.9882	0.9390	0.9333	0.9690	0.9610	0.9831	0.9651
ResNet50	0.9840	0.9828	0.9829	0.9759	0.9884	0.9699	0.9755	0.9795
VGG16	0.9840	0.9731	0.9771	0.9556	0.9884	0.9734	0.9718	0.9731
VGG19	0.9822	0.9720	0.9657	0.9574	0.9826	0.9876	0.9755	0.9734

Figure 4 also contains a comparison of the two evaluations detailed in the tables.

Study of similarity between D1 and D2 breeds In this second part, we will perform a similarity study between the breeds of both datasets to determine how similar the D2 breeds are to the D1 breeds. Figure 5 shows the proportion of D2 images classified as each of the classes known to the models. It is assumed that the more significant the proportion of images classified as a class, the greater the similarity between those two races.

Discussion An expert in the horse world assessed manually the degree of morphological similarity, with values in the range 0%-100%, of each pair of different

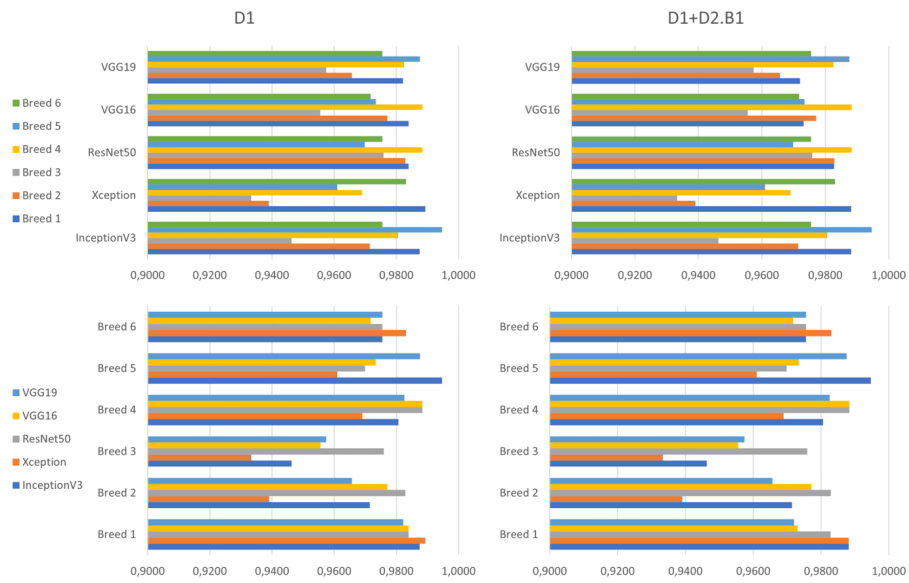


Fig. 4: Accuracy comparative between VGG16, VGG19, ResNet50, Exception and InceptionV3 for D1 and D2 breeds

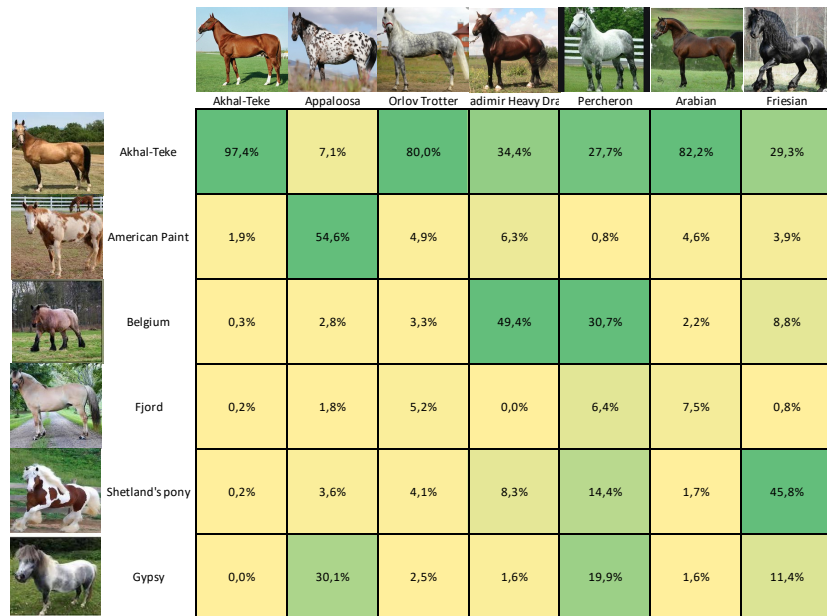
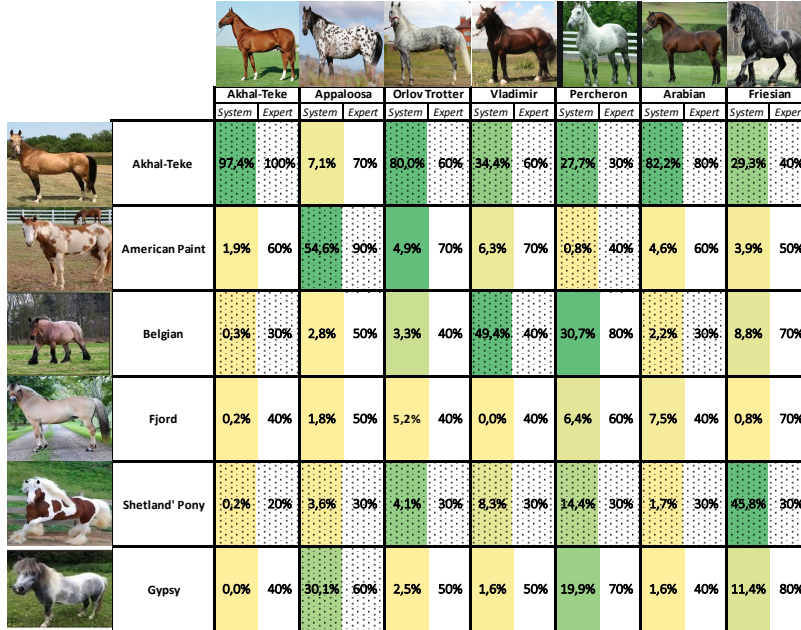


Fig. 5: Similarity between known and unknown breeds

breeds from both datasets. Figure 6 shows the figures of automatic similarity estimation (System) and expert assessment for each pair of breeds (Expert), where it can be remarked some agrees between the system and the human expert (in dotted pattern). The similarity estimated by the system agrees barely the expert for the Akhal-Teke/Shetland'pony compared to the D2 breeds. Moreover, in general terms almost the 50% of estimations agree.









		Akhal-Teke		Appaloosa		Orlov Trotter		Vladimir		Percheron		Arabian		Friesian	
		System	Expert	System	Expert	System	Expert	System	Expert	System	Expert	System	Expert	System	Expert
	Akhal-Teke	97,4%	100%	7,1%	70%	80,0%	60%	34,4%	60%	27,7%	30%	82,2%	80%	29,3%	40%
	American Paint	1,9%	60%	54,6%	90%	4,9%	70%	6,3%	70%	0,8%	40%	4,6%	60%	3,9%	50%
	Belgian	0,3%	30%	2,8%	50%	3,3%	40%	49,4%	40%	30,7%	80%	2,2%	30%	8,8%	70%
	Fjord	0,2%	40%	1,8%	50%	5,2%	40%	0,0%	40%	6,4%	60%	7,5%	40%	0,8%	70%
	Shetland' Pony	0,2%	20%	3,6%	30%	4,1%	30%	8,3%	30%	14,4%	30%	1,7%	30%	45,8%	30%
	Gypsy	0,0%	40%	30,1%	60%	2,5%	50%	1,6%	50%	19,9%	70%	1,6%	40%	11,4%	80%

Fig. 6: Comparative study of the breeds similarity assess by the the proposal versus an human expert

4 Conclusion and future work

We have developed an exhaustive analysis of five pre-trained models for horse-breeds images classification from two different datasets, D1 and D2. On the one hand, the obtained results for D1, outperformed clearly the baseline algorithm by Atabay[2], on the other hand, the models were used to compare images from different breeds obtaining similar results to the ones by an human expert.

For future work, it is proposed to continue with academic studies analyzing the performance of the architectures used for this study with experiments using D2 and new datasets composed by modifications of the images contained in D2.

Other interesting application of the transfer learning models can be the classification of breeds from isolated or combined parts of the horse body like the

head, legs, body, rump and tail for example. This way, it's possible selecting just some desired parts of the studs in the breeding process.

Acknowledgement

This research has been funded partially by Spanish Ministry of Economy, Industry and Competitiveness (MINECO) under grant TIN2017-84804-R/PID2020-112726RB-I00. In addition, we would like to thank Jose Sánchez Cebollada for their valuable assistance as expert on horse breeds morphology.

References

1. AFP, T.L.J.F.: French vineyards revive horse-drawn ploughs (2016), <https://www.thelocal.fr/20160814/picture-postcard-french-vineyards-revive-horse-drawn-ploughs/>
2. Atabay, H.: Deep Learning for Horse Breed Recognition. The CSI Journal on Computer Science and Engineering **15**(1), 45–51 (2017), https://www.researchgate.net/profile/Habibollah_Agh_Atabay3/publication/325946198_Deep_Learning_for_Horse_Breed_Recognition/links/5b2e0222aca2720785db33b3/Deep-Learning-for-Horse-Breed-Recognition.pdf
3. de Castelbajac, H.: Viticulture / oenology -Viticulture- : 3 winegrowers use horses (2020), <https://www.vitisphere.com/news-93104-3-of-French-winegrowers-use-horses.htm>
4. Chollet, F.: Xception: Deep Learning with Depthwise Separable Convolutions. CoRR **abs/1610.02357** (2016), <http://arxiv.org/abs/1610.02357>
5. de la Cal (University of Oviedo), E, García, E.U.o.O.: Ciencias de Datos en el Mundo Equino. In: II Congreso Internacional AINISE (2020), <https://www.ainise.org/ponentes/dr-enrique-de-la-cal/>
6. van Gemert, J.C., Verschoor, C.R., Mettes, P., Epema, K., Koh, L.P., Wich, S.: Nature conservation drones for automatic localization and counting of animals. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). vol. 8925, pp. 255–270. Springer Verlag (2015). https://doi.org/10.1007/978-3-319-16178-5_17
7. Hanot, P., Guintard, C., Lepetz, S., Cornette, R.: Identifying domestic horses, donkeys and hybrids from archaeological deposits: A 3D morphological investigation on skeletons. Journal of Archaeological Science **78**, 88–98 (feb 2017). <https://doi.org/10.1016/j.jas.2016.12.002>
8. Hanot, P., Herrel, A., Guintard, C., Cornette, R.: Morphological integration in the appendicular skeleton of two domestic taxa: The horse and donkey. Proceedings of the Royal Society B: Biological Sciences **284**(1864) (oct 2017). <https://doi.org/10.1098/rspb.2017.1241>, <https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2017.1241>
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. CoRR **abs/1512.03385** (2015), <http://arxiv.org/abs/1512.03385>
10. Kaggle: Kaggle.com (2020), <https://www.kaggle.com>

11. Merkies, K., Paraschou, G., McGreevy, P.D.: Morphometric Characteristics of the Skull in Horses and Donkeys—A Pilot Study. *Animals* **10**(6), 1002 (jun 2020). <https://doi.org/10.3390/ani10061002>, <https://www.mdpi.com/2076-2615/10/6/1002>
12. 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.: ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)* **115**(3), 211–252 (2015). <https://doi.org/10.1007/s11263-015-0816-y>
13. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. In: Bengio, Y., LeCun, Y. (eds.) 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings (2015), <http://arxiv.org/abs/1409.1556>
14. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the Inception Architecture for Computer Vision. *CoRR* **abs/1512.00567** (2015), <http://arxiv.org/abs/1512.00567>
15. Van Gemert, J.C., Verschoor, C.R., Mettes, P., Epema, K., Koh, L.P., Wich, S.: Nature Conservation Drones for Automatic Localization and Counting of Animals. Springer (2014). <https://doi.org/10.1007/978-3-319-16178-5>, <http://dutchuas.nl/http://idisnow.com>
16. Vayssade, J., Arquet, R., In, M.B.C., Electronics, undefined 2019: Automatic activity tracking of goats using drone camera. Elsevier (2019), <https://www.sciencedirect.com/science/article/pii/S0168169918312894>
17. www.wineterroirs.com: Wine Tasting, Vineyards, in France: Draft Horse in the vineyard (2010), https://www.wineterroirs.com/2010/04/draft_horse.html