Domain Transfer for Delving into Deep Networks Capacity to De-Abstract Art

Corneliu Florea[(✉)], Mihai Badea, Laura Florea, and Constantin Vertan

Image Processing and Analysis Laboratory,
Univeristy Politehnica of Bucharest, Bucharest, Romania
`corneliu.florea@upb.ro`

Abstract. Humans are capable of perceiving a natural scene at a glance and painters, through their representations, push this capacity to the limit by abstraction. Yet we still do it... What about machines? In this paper we address the problem of recognizing the theme (scene type) in digitized paintings. The approach is based on the Convolutional Neural Network framework and the chosen architecture is the recent Residual Network. In the first level of evaluation, we determine the recognition rate of a CNN given a database of 80,000 annotated digitized paintings. In the second level we evaluate the impact of extending the training database with photographs directly and via two domain adaptation functions and thus we are able to assess the abstraction level that CNN is capable to achieve.

1 Introduction

Pablo Picasso said "There is no abstract art. You must always start with something. Afterward you can remove all traces of reality." As art follows humanity through its entire history, if one integrates with respect to the level of abstraction he will note that the closer to the present moment we are, the more traces of reality have been removed.

In last period two trends favored the apparition of works similar to this one. First there were consistent efforts to digitize more and more paintings so that modern systems may learn from large databases. Two of such popular efforts are Your Paintings (now Art UK[1]) which contains more than 200.000 paintings and WikiArt[2] with around 100.000 paintings. The databases come with multiple annotations. For this work we are particulary interested in annotations about the painting's theme or scene type. From this point of view, a more complete database is the WikiArt collection, where the labelling category is named *genre*.

Secondly the development of the Deep Neural Networks allowed classification performance that was not imagined before. Here, we will inspire from the use of the more popular Convolutional Neural Networks (CNN). Given the achievable

---

C. Florea and M. Badea—Equally contributed.

[1] http://artuk.org/.

[2] http://www.wikiart.org/.

performance, the focus may switch from improving the performance to its use in practical tasks.

Starting from the idea that the Deep Neural Networks share similarities with the human vision [7] and the fact that such networks are already proven to do great jobs in other perception inspired areas like object recognition or even in creating artistic images, we ask ourselves if they can pass the abstraction limit and correctly recognize the scene type of a painting. We will first compare the results of residual network (ResNet) on the standard WikiArt database with previous methods from state of the art. We will then test different domain transfer augmentations to see if they can help increase the achieved recognition rate and also if the network is capable to pass the abstraction limit and learn from different types of images that contain the same type of scenes. Furthermore, we introduce several alternatives for domain transfer to achieve a dual-task: improve the scene recognition performance and understand the abstraction capabilities of machine learning systems.

Regarding the deep networks, multiple improvements have been proposed. In many situations, if given database is smaller, better performance is reachable if the network parameters are previously found for a different task on large database such as ImageNet. Next, these values are updated to a given task. This is called fine-tuning and it is a case of transfer learning. As our investigation is related to a different domain transfer we will avoid to use both so to establish clearer conclusions. To compensate, we are relying on the recent architecture of the residual networks (Resnet [16]) that was shown to be able to overcome the problem of vanishing gradients, reaching better accuracy for the same number of parameters when compared to previous architectures.

The remainder of the paper is organized as follows: Sect. 2 presents previous relevant work, Sect. 3 summarizes the CNN choices made and Sect. 4 will discuss different aspects of painting understanding. Section 5 presents the used databases, while implementation details and results are in Sect. 6. The paper ends with discussions about the impact of the results.

2   Related Work

Object and Scene Recognition in Paintings. Computer based painting analysis has been in the focus of the computer vision community for a long period. A summary of various directions approached, algorithms and results for not-so-recent solutions are in the review of Bentowska and Coddington [5]. However the majority addressed style (art movement) or artist recognition. Object recognition has been in the focus of Crowley and Zisserman [9] while searching through YourPaintings dataset with learning on photographic data.

Scene recognition in paintings is also named genre recognition following the labels from the WikiArt collection. This topic was approached by Condorovici et al. [8] and by Agarwal et al. [1]; both works, using the classical feature + classifier approach, tested smaller databases with few (5) classes: 500 images - [8] and 1500 images - [1]. More extensive evaluation, using data

from WikiArt, was performed by Saleh and Elgammal [22], which investigated an extensive list of visual features and metric learning to optimize the similarity measure between paintings and respectively by Tan et al. [25], which employed an AlexNet architecture [19] of CNN initialized (fine tuned) on ImageNet to recognize both style and genre of the painting.

The process of transferring knowledge from natural photography to art objects has also been previously addressed beyond the recent transfer from ImageNet to Wikiart [25]. 3D object reconstruction can be augmented if information from old paintings is available [3]. Classifiers (deep CNNs) trained on real data are able to locate objects such as cars, cows and cathedrals [10]. The problem of detecting/recognizing objects in any type of data regardless if it is real or artistic was named cross-depiction by Hall et al. [15]; however the problem is noted as being particular difficult and even in the light of dedicated benchmarks [6], as the results show a lot of place for improvement. Another comment is that all solutions that showed some degree of success did it for older artistic movements where scene depiction was without particular abstraction. To our best knowledge there isn't any significant success for more modern art.

Of particular interest for our work is the algorithm recently introduced by Gatys et al. [14], which used various layers of CNN trained for object recognition to separate the content from the style of an image and to enable style transfer between pairs; the most impressive results are in the transfer of artistic style to photographs rendering the later as being painted in rather abstract ways.

Scene Recognition in Photographs. Scene recognition in natural images is an intensively studied topic, but under the auspices of being significantly more difficult than object recognition or image classification [30]. We will refer the reader to a recent work [17] for the latest results on the topic. We merely note that the introduction of the SUN database [27] (followed by expansions) placed a significant landmark (and benchmark) on the issue and that it was shown that using domain transfer (e.g. from Places database), the performance may be improved [30].

Scene Recognition by Humans. While it is outside the purpose of this paper to discuss detailed aspects of the human neuro-mechanisms involved in scene recognition, following the integrating work of Sewards [23] we stress one aspect: compared to object recognition where localized structures are used, for scene recognition the process is significantly more tedious and complex. Object recognition, "is solved in the brain via a cascade of reflexive, largely feedforward computations that culminate in a powerful neuronal representation in the inferior temporal cortex" [11]. In contrast, scene recognition includes numerous and complex areas as the process starts with peripherical object recognition, continues with central object recognition, activating areas such entorhinal cortex, hippocampus and subiculum [23].

Concluding, there is consensus, from both neuro-scientist and from the computer vision community that scene recognition is a particularly difficult task.

This task becomes even harder when the subject images are heavily abstracted paintings from modern art.

3   Architecture and Training

In the remainder of the paper, we will use the Residual Network (ResNet) [16] architecture with 34 layers. All the hyper-parameters and the training procedure follows precisely the original ResNet [16]. Nominally, the optimization algorithm is 1-bit Stochastic Gradient Descent, the initialization is random (i.e. from scratch) and when the recognition accuracy on the validation set plateaus we decrease the learning rate by a factor of 10. The implementation is based on the CNTK library[3].

Database Augmentation. To improve the recognition performance various database augmentation scenarios have been tested. The ones that have produced positive effects are flipping and slight rotation. The flipped samples are all horizontal flips of the original images. Regarding the rotations, all images have been rotated either clockwise or counterclockwise with 3°, 6°, 9° or 12°. We do not refer here to the domain transfer experiments.

4   Painting Understanding and Domain Transfer

A task that we undertake is to get an understanding of the machine learning systems (in our case deep CNN) grasp of art. For CNN, the favorite visualization tool has been proposed by Zeiler and Fergus [28] by introducing deconvolutional layers and visualizing activations maps onto features.

Attempts to visualize CNNs for scene recognition using this technique indicated that activation is related to objects, thus leading to the conclusion that multiple object detectors are incorporated in such a deep architecture [29]. In parallel, visualization of activation for genre [25] shown that, for instance, the landscape type of scene leads to activating almost the entire image, thus being less neat to draw any conclusion. Consequently, we tried a different approach to investigate the intrinsically mechanisms of deep CNN. Our approach exploits domain transfer.

Given the increased power of machine learning systems and the limited amount of data available to a specific task, a plethora of transfer learning techniques appeared [20]. The process of transfer learning is particular popular when associated with deep learning. First, let us recall that the lower layers of deep nets trained on large databases are extremely powerful features when coupled with powerful classifier (such as SVM) and maybe a feature selector, no matter the task [13]. Secondly, the process of *fine tuning* deep networks assumes taking a network that has been pre-trained on another database, and, using a small learning rate, only adapt it to the current task.

---

[3] Available at https://github.com/microsoft/cntk/.

In contrast, the concept of domain transfer or domain adaptation appeared as an alternative to increase the amount of information over which a learner may be trained directly (without fine tuning) in order to improve its prediction capabilities. Many previous solutions and alternatives have been introduced. We will refer to the work of Ben-David et al. [4] for theoretical insights on the process.

It has been shown that domain transfer is feasible and the resulting learner has improved performance if the two domains are adapted. Saenko et al. [21] showed that using a trained transformation, the domain transfer is beneficial. We investigate two alternatives. Firstly, we consider the Laplacian style transfer introduced by Aubry et al. [2]. They use a variant of bilateral filter to transfer the edginess from the reference, artistic image to the realistic photo. Secondly we consider the neural algorithm introduced by Gatys et al. [14]. Using a deep CNN they decompose an image into style and content. Intuitively the major difference between an artistic image and a photo is the style; doing style transfer, the second will be adapted to the first one's domain. Among several deep CNN architectures investigated, in the original work and confirmed by our experiments, only the VGG19 [24] leads to qualitative results.

5   Databases

For the various experiments undertaken, two databases have been employed. These are the WikiArt paintings dataset which was collected from Internet and used for the first time in this shape by Karayev et al. [18] and respectively the SUN database [27]. The former contains the bulk of the images used for training and testing, while the latter is used only as an auxiliary database for domain transfer experiments.

5.1   WikiArt Database

The WikiArt database contains approximately 80,000 digitized images of fine-art paintings. They are labelled within 27 different styles (cubism, rococo, realism, fauvism, etc.), 45 different genres (illustration, nude, abstract, portrait, landscape, marina, religious, literary, etc.) and belong to more than 1000 artists. To our knowledge this is the largest database currently available that contain genre annotations. Due to the fact that some classes are limited as number of examples, we chose to use only the ones that are well illustrated.

For our tests we considered a set that contains 79434 images of paintings. Some of the scene types were not well represented (i.e. less than 200 images) and we gathered them into a new class called "Others". This led to a division of the database into 26 classes. The names of the classes and the number of training and testing images in each class can be seen in Table 1.

We note that annotation is weak, as one may find arguable labels. For instance "literary" and "illustration" categories may in fact have "landscape" themes. Another observation is that there exists two "collector" classes: "others"

and "genre". However, as this distribution matches practical situations, we used the database as it is, without altering the annotations.

5.2   Natural Scene Databases

As an additional source of real data we have relied on images from the SUN database [27]. In its original form, it contains 899 classes and more than 130,000 images. Yet only a few classes, which have a direct match with our database were selected. These classes and the number of images in each class added in the training process can be seen as the top, green segment in Fig. 1.
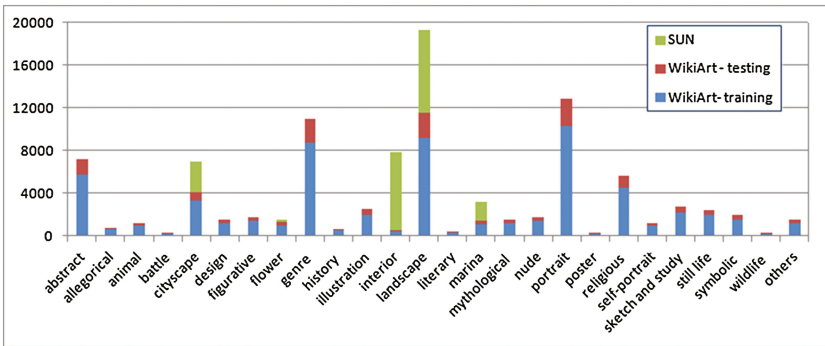

Fig. 1. Structure of the used databases

Table 1. Comparison with state of the art methods. The table is horizontally split to group solutions that have used databases with similar size. Acronyms: BOW - Bag of Words, ITML - or Iterative Metric Learning; pHoG - HoG pyramid as in [12]; pLBP - LBP pyramid implemented in VLFeat [26] DeCAF [13] assumes the first 7 levels of AlexNet trained on ImageNet.

Method	No. classes	No. images	Test ratio	Accuracy (%)
Agarwal et al. [1] - SIFT + BOW	5	1500	10%	82.53
Agarwal et al. [1] - ensemble				84.46
Saleh and Elgammal [22] - Classemes + Boost	10	63.691	33%	57.87
Saleh and Elgammal [22] - Classemes + ITML				60.28
Saleh and Elgammal [22] - Classemes + Fusion				60.28
Tan et al. [25] AlexNet - scratch			n/a	69.29
Tan et al. [25] CNN- finetune				**74.14**
ResNet 34 - scratch			20%	*73.74*
pHoG + SVM	26	79,434	20%	44.37
pLBP + SVM				39.58
DeCAF + SVM				59.05
AlexNet - scratch				53.02
ResNet 34 - scratch				**61.15**

6   Implementation and Results

It was our main interest to study the various ways in which the performance of the different systems tested can be increased. This included experiments on the classification methods themselves or various alterations brought to the database.

6.1   Comparison with State of the Art Methods

Agarwal et al. [1], Tan et al. [25] and Saleh and Elgammal [22] used the WikiArt database for training and testing in order to classify paintings into different genres. While the first used a very small subset, the later two cases focused on 10 classes from the database (Abstract, Cityscape, Genre, Illustration, Landscape, Nude, Portrait, Religious, Sketch and Study and Still life) leading to ∼63 K images.

We have adopted the division from Karayev et al. [18] as working with a complete version of Wikiart. Furthermore, we stress that in our case the images from training and testing are completely different and are randomly chosen.

In order to compare our results to the ones reported by the mentioned articles, we selected the same classes of paintings for training and testing. While in a case [22], the test-to-train ratio is mentioned, in the second, [25], it is not. Under these circumstances the comparison with prior art is, maybe, less accurate.

The results (showed in Table 1) indicates that the proposed method gives similar results with previous [25] with the difference that they use a smaller network (AlexNet) but fine–tuned, while we have used a larger one, but initialized from scratch. Furthermore, we report the average over 5 runs.
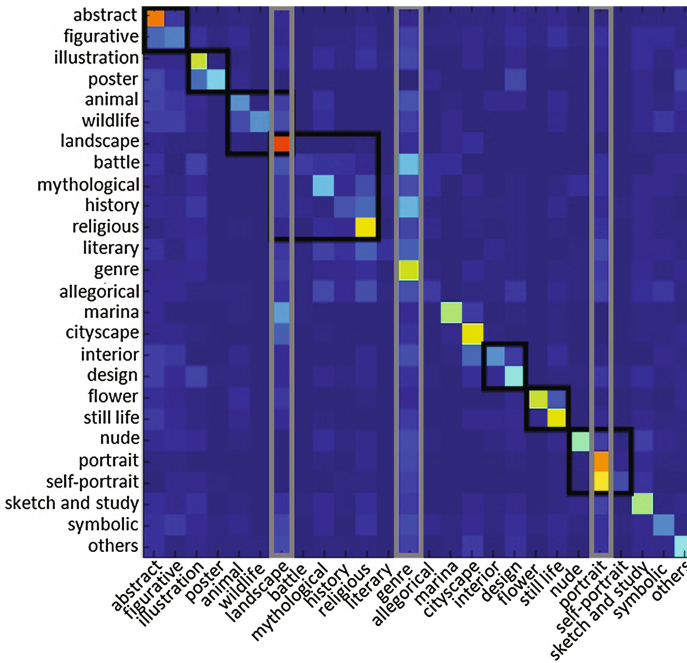
6.2   Confusion Matrix

Visual examples of paintings are shown in Fig. 4. The confusion matrix for the best performer on the 26-class experiment is in Fig. 2. We have marked classes that are particular confused. It should be noted that from a human point of view, there is certain confusion between similar genres such as historical–battle–religious, portrait–self portrait, poster–illustration, animal–wildlife etc. Some of these confusable images are, in fact, shown in Fig. 4. Consequently we argue that the top-5 error is also relevant, as in many cases there are multiple genre labels that can be truthfully associated with one image. For the best proposed alternative, ResNet with 34 layers the Top-5 error is 11.85% - corresponding to a **88.15**% accuracy. For the 10-class experiment the top–5 accuracy is **96.75**%.

6.3   Additional Experiments

For the following experiments we will refer solely to the 26 classes test as it the most complete.

As many experiments require a significant amount of time, we restrained the training to 125 epochs. In this case the training takes ∼20 h on NVidia GeForce

Fig. 2. The confusion matrix for the 26 class with 300 epochs training variant.

980 TI compared to 55 h for the 300 epochs alternative at the expense of 2% accuracy (Fig. 3).

Stochastic Effect. The first test studied the effect of the stochastic nature of deep neural networks. Factors such as the random initialization of all the

| Self-Portrait | Portrait | Interior | Design | Poster |
| History | Battle | Genre | Nude |
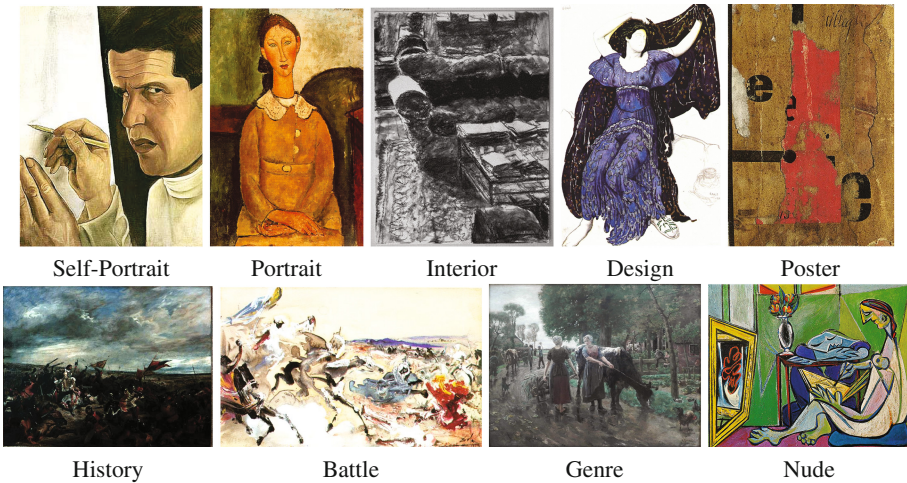
Fig. 3. Examples of the genres as illustrated in the database. Please note that other genre labels may be used for each image, thus arguing for the use of top–5 classification.

parameters can influence the results of any considered network. We ran the ResNet 34 several times on 26 classes and the accuracy results had a mean of 59.1% (top–1 accuracy) and a standard deviation of 0.33%. The results underline the fact that even though there is some variation caused by randomness, it does not influence the system significantly.

Influence Artistic Style. Prior art [29] suggested that even the case of the scene, in fact a deep network builds object detectors and it can recognize objects it has seen before. To study this aspect we devise the following experiment. Given the full 79 K images genre database, we selected all the images that are associated with Cubist and Naive Art styles and placed them in testing, resulting in 4,132 images for evaluation and ∼75 K for training. While numerically this is a weaker test than previous, the results are considerably worse (50.82% top–1 accuracy and 82.10% top–5). This is due to the fact that these particular styles are rather different from the rest and the learner had no similar examples in the training database. Also these results argues for *a style oriented domain adaptation*.

(a)   (b)   (c)   (d)
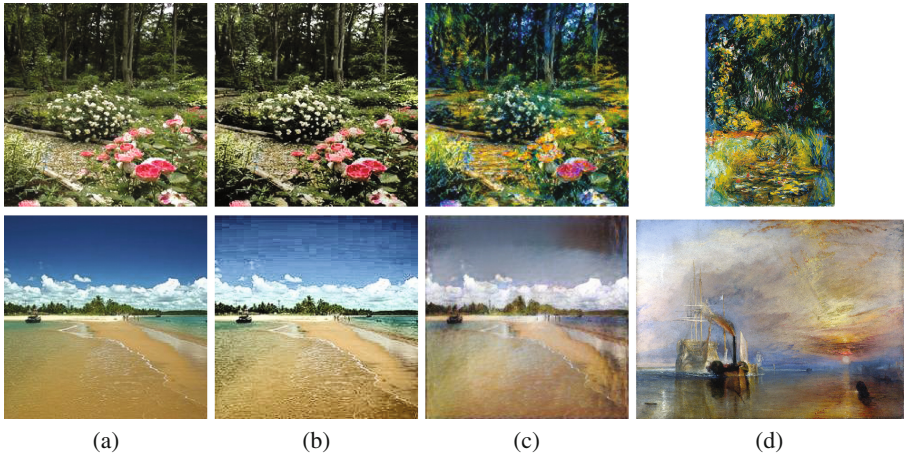
Fig. 4. Illustration of the domain transfer and domain adaptation experiments. Column (b) and (c) marks the image transformed. Columns are: (a) original photograph; (b) the image obtained after the Laplacian transfer [2]; (c) the image obtained after the neural transfer [14]; (d) the reference painting.

6.4   Domain Transfer

The domain transfer experiments consisted of separately augmenting certain classes with examples from the SUN database. Table 2 contains both overall performance for each class augmentation and the change brought to the concerned class (shown in the "Modifier" column). This measure takes into account the value of correct classifications of the regular networks, rather than number of existing samples. The experiment assumes adding each class separately. The

Table 2. The effect of adding extra samples from the SUN database. The "Acc" and "Acc-5" refers to the overall set accuracy when only the first and the top–5 results are taken into account. "Modif" refers to the improvement of the particular class.

Transfer	Original			Laplacian - [2]			Neural - [14]		
Class	Acc [%]	Acc-5 [%]	Modif. [%]	Acc [%]	Acc-5 [%]	Modif. [%]	Acc [%]	Acc-5 [%]	Modif. [%]
Cityscape	58.8	88.37	0.415	58.95	87.59	0.53	n/a	n/a	n/a
Flower painting	58.66	87.23	0.7	59.25	87.89	1.59	57.21	86.9	3.97
Interior	58.33	87.73	17.54	58.30	87.46	3.78	n/a	n/a	n/a
Landscape	59.32	88.43	3.85	58.61	88.13	0.18	n/a	n/a	n/a
Marina	58.35	87.78	15.7	59.50	88.13	1.55	57.22	87.33	1.16

neural style transfer method is very slow requiring 10–30 min to adapt an image. Thus we restrict to augmenting only Marina and Flower Painting classes.

Adding all transferred (adapted) images produced a similar effect, the variation being smaller than stochastic variance (overall accuracy of 59.05%).

As one may notice the overall results are not conclusive, variation being in the stochastic variance. However each of the transfers augmented the classification on the respective class. The most visible numerical effect over the entire database is obtained by adding original interior images; this is the only class where the number of added images is significantly larger than the associated paintings. Also given the transfer, the improvement is associated with styles such academism or realism, that contain very realistic rendering of the original scene, without much abstraction.

The images produced by the Laplacian transfer, while they look more abstract, do not seem "painted"; this domain adaptation does not improve the objective evaluation. The transfer, here, focuses on local contrast and grayscale dynamic range, while CNNs are related to structure. As shown in Fig. 4(b), there is no impression of painting in the images produced, thus it is hardly related with testing examples and is unhelpful when partitioning the data space.

We have found to be somehow surprising that, although the neural style produce images, which seem similar to a painting, its numerical effect is not as dramatic as we expected. However, we believe that the explanation is given by the quantity. The process is lengthy and we only added a small number of images, that are not able to actually fill the data space so that the CNN is able to draw rigorous borders. To see if this is the case we devised two experiments where the number of transferred images is comparable with the one from the standard database.

Neurally Transferring few Images on a Small Database. For these experiments, we have produced, using the neural style transfer algorithm [14], images for three genres: "cityscape" – 262; "flower paintings" – 180 and "marina" – 229. We have considered the case with 26 classes.

For the first experiment, we aimed to see what happens if, for the chosen classes, instead of paintings we provide mostly transferred images. The results are illustrated in Table 3. Initially, we have removed completely any training data for the three classes; obviously there was no correct recognition for these classes.

Than, we have added only the transferred images. In this case there are correct recognitions, even for paintings in abstract style, while quite few, showing that the neural transfer may help and provides relevant data. Afterwards we have added iteratively more images and we noted increasing recognitions. At the end we have removed the transferred images and noted a decrease, again confirming the beneficial effect of domain adaptation.

Another observation related to results is that the impact on the "flower paintings" genre is much reduced when compared to other two, "cityscape" and "marina". A possible explanations is related to the content: for paintings, the "flower paintings" genres refers typically to still flowers in a vase; in contrast for images, "flowers" are from garden, occupying much smaller areas from the image. Such an example is in Fig. 4.

For the second experiment we have reduce the contribution of each class to a number comparable with those transferred (namely to 250). The numerical results are showed in lower part of Table 3. One may see that if the quantity of neural transferred paintings is comparable with original data and the content of the two sets is similar (i.e. for cityscape and marina it is, while for flower is much less), the transfer is again beneficial. Thus, this experiments also shows that the neural style transfer may act as a domain adaptation function.

7   Discussion

In this paper we discuss the CNN capabilities to recognize the scene in a painting. A first contribution is that we clearly showed that machine learning systems

Table 3. Painting recognition accuracy, when classes that can be augmented by neural style transfer [14] have a diminished number of original paintings. The classes of interest are "cityscape", "flower paintings", "marina". Number of images transferred: "cityscape" – 262; "flower paintings" – 180 and "marina" – 229. Paintings used for training in other classes are 57363. For the second experiment, we have considered 250 paintings per class, totalling for 23 classes 5750.

Exp. No.	Paint. in other classes	Paints per interest class	Transferred images	Recognized images			
				Cityscape (From 764)	Flower (From 252)	Marina (From 259)	All classes (From 15708)
1	all	0	0	0	0	0	8024
	all	0	all	37	0	6	8084
	all	50	all	29	17	5	8255
	all	200	all	107	67	24	8364
	all	250	all	149	73	68	8584
	all	250	0	121	48	59	8328
2	5750	0	all	125	24	48	3354
	5750	250	0	219	91	120	3357
	5750	250	all	287	141	145	3561

(deep CNN) are confused by the abstraction level from art. The experiment with abstract art showed that they cannot easily generalize with respect to style.

Furthermore we experimented with domain transfer as an alternative to increase the overall performance and we have found that: (1) sheer numbers of photographs have a beneficial effect by improving the performance over styles with realistic depictions; (2) the CNNs are confused by artistic rendering (i.e. artistic style); (3) the neural transfer style may act as domain transfer adaptation. However significant increase into artistic scene recognition is opposed by the large duration of the neural style transfer. Thus speed-ups of the latter are highly necessary.

References

1. Agarwal, S., Karnick, H., Pant, N., Patel, U.: Genre and style based painting classification. In: WACV, pp. 588–594 (2015)
2. Aubry, M., Paris, S., Hasinoff, S.W., Kautz, J., Durand, F.: Fast local Laplacian filters: theory and applications. ACM Trans. Graph. **33**(5), 167 (2014)
3. Aubry, M., Russell, B., Sivic, J.: Painting-to-3D model alignment via discriminative visual elements. ACM Trans. Graph. **33**(2), 14 (2013)
4. Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., Wortman Vaughan, J.: A theory of learning from different domains. Mach. Learn. J. **79**(1–2), 1–2 (2010)
5. Bentkowska-Kafel, A., Coddington, J.: Computer vision and image analysis of art. In: SPIE (2010)
6. Cai, H., Wu, Q., Hall, P.: Beyond photo-domain object recognition: Benchmarks for the cross-depiction problem. In: ICCV workshops - TASK, pp. 262–273 (2015)
7. Cichy, R., Khosla, A., Pantazis, D., Torralba, A., Oliva, A.: Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. Scientific reports 6 (2016)
8. Condorovici, R.G., Florea, C., Vertan, C.: Painting scene recognition using homogenous shapes. In: Blanc-Talon, J., Kasinski, A., Philips, W., Popescu, D., Scheunders, P. (eds.) ACIVS 2013. LNCS, vol. 8192, pp. 262–273. Springer, Cham (2013). doi:10.1007/978-3-319-02895-8_24
9. Crowley, E., Zisserman, A.: The state of the art: object retrieval in paintings using discriminative regions. In: BMVC (2014)
10. Crowley, E.J., Zisserman, A.: The art of detection. In: Hua, G., Jégou, H. (eds.) ECCV 2016. LNCS, vol. 9913, pp. 721–737. Springer, Cham (2016). doi:10.1007/978-3-319-46604-0_50
11. DiCarlo, J., Zoccolan, D., Rust, N.: How does the brain solve visual object recognition? Neuron **73**(3), 415–434 (2012)
12. Dollar, P., Appel, R., Belongie, S., Perona, P.: Fast feature pyramids for object detection. T. PAMI **36**(8), 1532–1545 (2014)
13. Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., Darrell, T.: DeCAF: a deep convolutional activation feature for generic visual recognition. In: ICML (2014)

14. Gatys, L., Ecker, A., Bethge, M.: A neural algorithm of artistic style. In: CVPR (2016)
15. Hall, P., Cai, H., Wu, Q., Corradi, T.: Cross-depiction problem: recognition and synthesis of photographs and artwork. Comput. Visual Media **1**(2), 91–103 (2015)
16. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR (2016)
17. Herranz, L., Jiang, S., Li, X.: Scene recognition with CNNS: objects, scales and dataset bias. In: CVPR (2016)
18. Karayev, S., Trentacoste, M., Han, H., Agarwala, A., Darrell, T., Hertzmann, A., Winnemoeller, H.: Recognizing image style. In: BMVC (2014)
19. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (eds.) Advances in Neural Information Processing Systems 25. pp. 1097–1105 (2012)
20. Lu, J., Behbood, V., Hao, P., Zuo, H., Xue, S., Zhang, G.: Transfer learning using computational intelligence: a survey. Knowl.-Based Syst. **80**, 14–23 (2015)
21. Saenko, K., Kulis, B., Fritz, M., Darrell, T.: Adapting visual category models to new domains. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6314, pp. 213–226. Springer, Heidelberg (2010). doi:10.1007/978-3-642-15561-1_16
22. Saleh, B., Elgammal, A.: Large-scale classification of fine-art paintings: learning the right metric on the right feature. In: International Conference on Data Mining Workshops. IEEE (2015)
23. Sewards, T.V.: Neural structures and mechanisms involved in scene recognition: a review and interpretation. Neuropsychologia **49**(3), 277–298 (2011)
24. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: ECCV (2014)
25. Tan, W.R., Chan, C.S., Aguirre, H.E., Tanaka, K.: Ceci n'est pas une pipe: a deep convolutional network for fine-art paintings classification. In: ICIP (2016)
26. Vedaldi, A., Fulkerson, B.: VLFeat: an open and portable library of computer vision algorithms. In: ACM MM, pp. 1469–1472 (2010)
27. Xiao, J., Hays, J., Ehinger, K.A., Oliva, A., Torralba, A.: Sun database: large-scale scene recognition from abbey to zoo. In: CVPR, pp. 3485–3492 (2010)
28. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8689, pp. 818–833. Springer, Cham (2014). doi:10.1007/978-3-319-10590-1_53
29. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Object detectors emerge in deep scene CNNS. In: ICLR (2015)
30. Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., Oliva, A.: Learning deep features for scene recognition using places database. In: NIPS (2014)