How Convolutional Neural Networks Remember Art

Eva Cetinic*, Tomislav Lipic*, Sonja Grgic†
*Rudjer Boskovic Institute, Bijenicka cesta 54, 10000 Zagreb, Croatia
†University of Zagreb, Faculty of Electrical Engineering and Computing, Unska 3, 10000 Zagreb, Croatia
ecetinic@irb.hr

***Abstract*—Inspired by the successful performance of Convolutional Neural Networks (CNN) in automatically predicting complex image properties such as memorability, in this work we explore their transferability to the domain of art images. We employ a CNN model trained to predict memorability scores of natural images to explore the memorability of artworks belonging to different genres and styles. Our experiments show that nude painting and portrait are the most memorable genres, while landscape and marine painting are the least memorable. Regarding image style, we show that abstract styles tend to be more memorable than figurative. Additionally, on the subset of abstract images, we explore the relation between memorability and other features related to composition and color, as well as the sentiment evoked by the image. We show that there is no correlation between symmetry and memorability, however memorability positively correlates with the image's probability of evoking positive sentiment.**

***Index Terms*—Image Memorability; Fine Art; Convolutional Neural Networks**

I. Introduction

Alongside successful employment of CNNs for various computer vision tasks, arises the need to better understand learned features, as well as their transferability across different domains. CNNs trained for a domain-specific tasks represent a valuable source for transferring knowledge to other domains and applications, especially when collecting ground-truth labels is laborious and expensive. One such case is estimating image memorability – the concept referring to how easy it is for a person to remember an image. It has been shown that people share the tendency to remember the same images [1], which indicates that the phenomenon of memorability has a universal nature and lies beyond our subjective experience. This also implies that certain image attributes are more memorable than others.

In order to better understand which visual properties contribute to the memorability of an image, we investigate the memorability of art images, particularly in relation to artistic genre and artistic style. The artistic genre is related with the traditional division of paintings based on the type of content depicted (e.g. portrait, landscape, still life, etc.). Style is a historically and contextually dependent concept, often considered to be more associated with specific visual properties such as color palette, composition, level of detail, brushwork, etc. In our work we explore the memorability of images belonging to specific genres and styles in order to investigate the relation of memorability with the content and distinctive visual characteristics of artworks. We use a CNN model trained to predict memorability scores of photographs [2] in order to obtain memorability scores of paintings. In this regard, besides investigating the memorability of artworks, we also question the transferability of memorability-related CNN features from the domain of natural images to art images.

The next section provides an overview of related work. The third section describes the experimental setup, particularly the dataset and the CNN model. The results presenting memorability scores for different genres and styles are demonstrated in the fourth section, together with the analysis of the correlation between memorability and other image features. The conclusion is given in the final section.

II. Related Work

Fine art images are a valuable source of historically relevant, as well as perceptually interesting visual information. The increasing emergence of digitized fine art collections facilitated various interdisciplinary research questions which include the use of different computer vision techniques. Most of the studies dealing with art image data, addressed the challenge of classifying paintings by artist [3, 4], style [5–8] or genre [9, 10] by exploring different image features and machine learning methods. Apart from classification, other topics of interest were addressed such as exploring influential connections among artists [11], recognizing objects in paintings [12] or computationally exploring the aesthetics of artworks [13]. However, to the best of our knowledge, the question of predicting memorability of artworks has not yet been systematically explored.

Image memorability has been studied by psychologists for a long time. However, recently it became a subject of interest within the computer vision community when Isola et al. introduced a framework for predicting image memorability based on global image descriptors [1, 14, 15]. They built a memorability labeled dataset by collecting human responses through a visual memory game and trained a support vector regression (SVR) model to map image features such as GIST, SIFT, HOG and pixel color histograms into memorability scores. They also analyzed the correlation between specific image features and memorability concluding that simple image features do not correlate strongly with memorability. However, they showed that content plays an important role in memorability, with photos of people being more memorable than photos of landscapes. Following their work, various other approaches have been proposed in order to improve memorability prediction by exploring different image features [16–

978-1-5386-6979-2/18/$31.00 ©2018 European Union

19]. A comprehensive overview of studies related to image memorability is given in [20].

The adoption of CNNs for the task of image memorability was introduced by Khosla et. al [2]. They built a very large dataset with 60,000 images annotated with human memory scores conducted through a memory game experiment. Using this large dataset, they fine-tuned a CNN pre-trained for object and scene recognition to create the MemNet model. The model achieved state-of-the-art memorability prediction performance, reaching rank correlation of 0.64, with 0.68 being the human consistency rank correlation. As this model is used in our study, we describe it in more detail in the following section.

III. Experimental setup

A. WikiArt data

The data used in this study consists of images obtained from WikiArt.org, a large publicly available collection of digitized paintings. The WikiArt collection contains images annotated with a large set of labels such as artist, style, genre, technique, etc. At the time of our data collection process, it contained 133220 artworks in total. However, for the purpose of this study we prepared two sub-collections, one with images of artworks belonging to different genres and the other with images associated with different styles. Based on different genre categories and number of images included per category, we build a genre dataset consisting of 60260 images belonging to 10 different genres, where each category is represented with at least 1300 images. Similarly, the style dataset consists of 67200 images belonging to 27 different style categories, with at least 500 images per class. All images were resized to 256 x 256 pixels.

B. MemNet model

The convolutional neural network MemNet [2] used in our experiments is a fine-tuned version of the Hybrid-CNN model [21]. The Hybrid-CNN architecture is based on CaffeNet [22], a slightly modified version of the AlexNet model [23] created by switching the ordering of pooling and normalization layers. It contains five convolutional layers and three fully connected layers. Max-pooling layers follow after the first, second and fifth layer, while dropout is implemented after the first two fully connected layers. The activation function for all weight layers is the rectification linear unit (ReLU). The input of the network is a 227 x 227 crop of the resized RGB image. Hybrid-CNN is trained to classify categories of objects and scenes based on a training set consisting of 3.5 million images from 1183 categories, obtained by combining the Places [21] and ImageNet [24] datasets. The MemNet model is created by fine-tuning the Hybrid-CNN using the LaMem dataset, a large dataset annotated with human memory scores conducted through a memory game experiment using Amazon Mechanical Turk. The memorability score is a single real value in the range [0,1]. Therefore the final softmax layer is replaced with the Euclidean loss layer prior to fine-tuning the Hybrid-Net for memorability prediction.

978-1-5386-6979-2/18/$31.00 ©2018 European Union

IV. Results

This section includes the experimental results and systematic overview of the memorability scores for genre and style categories. Furthermore, it includes an analysis of the correlation between memorability and several different low and high level image features.

A. Genre memorability

To explore how memorable are particular artistic genres, we feed the images from the genre dataset to MemNet in order to obtain predictions of the memorability score for each image. Based on this score, we calculate the average memorability for each genre by dividing the sum of image-specific memorabilities with the number of images in each genre category. The nude painting category has the highest average memorability score of 0.853, while the lowest score of 0.656 is obtained for the landscape category. The box plot in Figure 1 shows the distribution of memorability scores across genres. The boxes are ordered by mean memorability score (marked with a red dot).

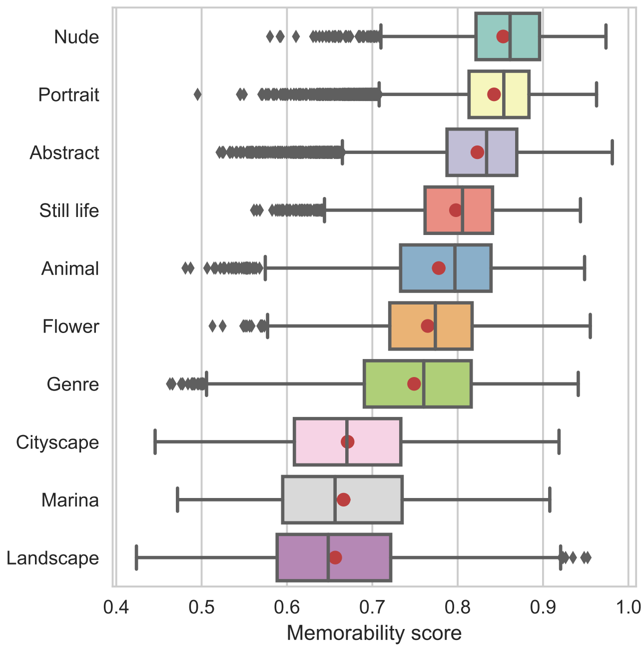

Fig. 1. Distribution of image memorability scores across genres.

Isola et al. [15] have shown that pictures with people tend to be more memorable than natural landscapes. The fact that nude paintings and portraits have the highest average memorability score, while landscape and marina paintings have the lowest score, shows that there is a consistency between art images and photographs when the subject of depiction is considered. Furthermore, abstract images tend to have a high memorability score. This can be seen in Figure 2 where the five most memorable (top row) and five least memorable (bottom row) from the genre subset are shown.

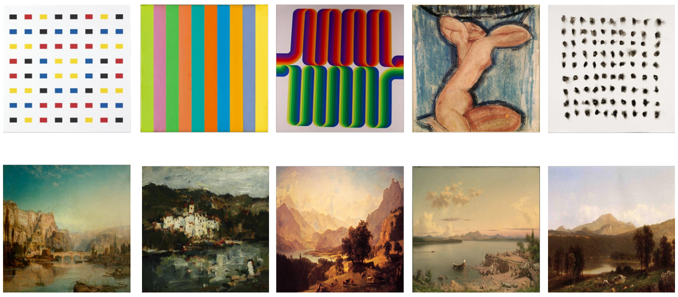

Fig. 2. Five most memorable (top row) and five least memorable (bottom row) artworks.

B. Style memorability

Similarly as in the case of genre memorability, we calculate average memorability scores for each artistic style. The box plot in Figure 3 shows the distribution of memorability scores across styles ordered by mean memorability.

The highest mean memorability score is obtained for abstract styles, while impressionism and romanticism tend to have a lower memorability score. Having in mind that impressionism and romanticism predominantly include landscapes paintings, the low memorability can be linked to the subject of depiction and the fact that landscapes are less memorable. However, a high memorability score of abstract styles shows that the absence of clearly defined object content tends to contribute to the increase of image memorability. This might be due to the fact that abstract images rarely appear in our daily visual experiences. Furthermore, because of the lack of recognizable object or scene content, the memorability of abstract images depends primarily on their visual image properties. This makes the subset of abstract images an interesting source for exploring the relation of visual features and memorability.

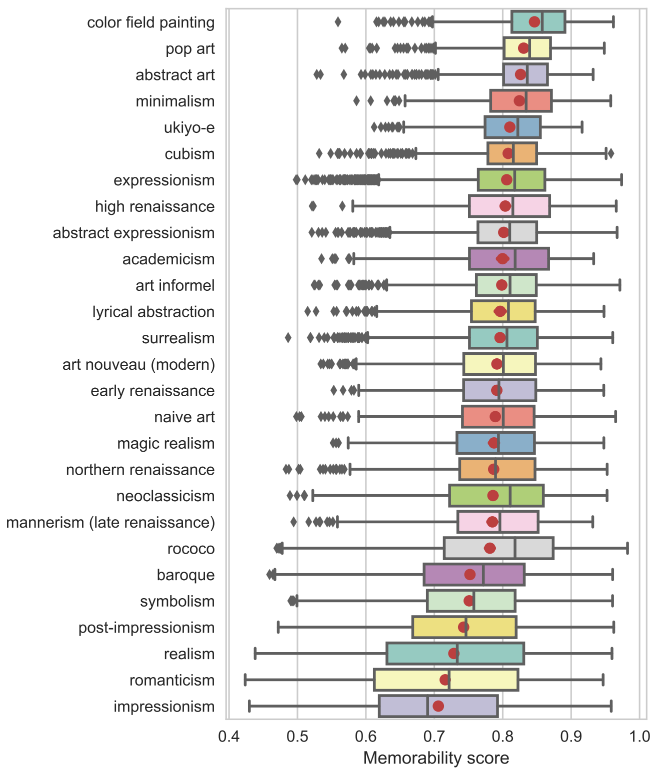

Fig. 3. Distribution of image memorability scores across styles.

C. Memorability in relation to other image features

In order to better understand how image properties correlate with memorability when recognizable object or scene content is omitted, we analyse different features from the style subset with the highest memorability score. We extract the 4096-dimensional CNN-features from the penultimate fully connected layer (fc7) of MemNet for 947 color field painting images and compute t-SNE [25] embedding. Based on the 2-dimensional embedding that respects the high-dimensional distances, we create a vizualization that shows how images are clustered according to MemNet learned features (see Figure 4). In Figure 4 we can observe that images are clustered according to some specific visual properties such as composition and color. However, in order to systematically explore the relation of memorability, we calculate the Spearman's rank correlation coefficient between memorability and several different low and high level features extracted from images included in the color field painting subset.

Regarding the composition of images, we calculate the ratios of horizontal and vertical edges in the image after detecting edges with the Sobel operator. Based on the calculation of Spearman's correlation coefficient, we conclude that

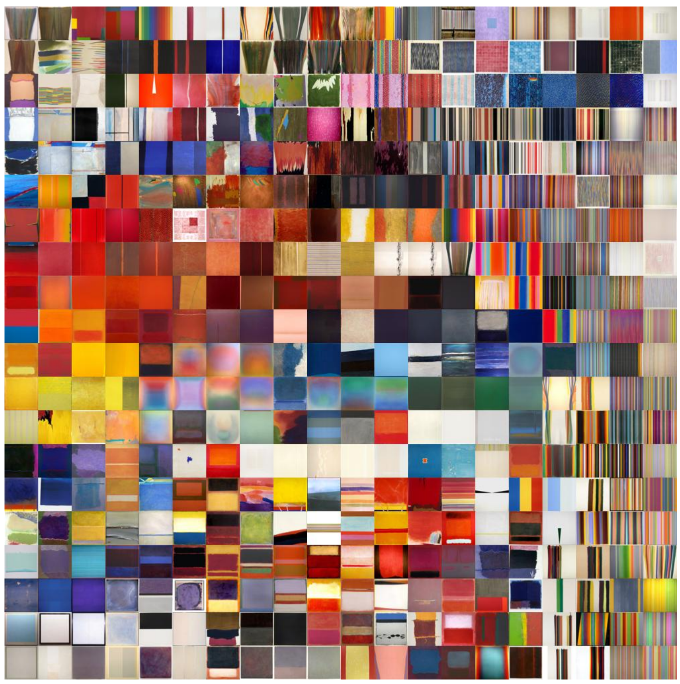

Fig. 4. t-SNE vizualization of MemNet fc7 features for color field painting images.

978-1-5386-6979-2/18/$31.00 ©2018 European Union

the verticality of composition does not correlate significantly with memorability. However, images with a high frequency of horizontal edges show a weak negative correlation with memorability. Furthermore, we calculate the vertical symmetry of images based on a measure proposed by Brachman and Redies in [26]. They define a new method for measuring symmetry in images by using CNN filter responses. The output of the symmetry measure is a value between 0 and 1, where higher values corresponding to more symmetrical images. This measure can be calculated from outputs of different CNN layers and for our purpose we use the activations of the fifth convolutional layer (conv5) with a patch size of 6. However, we show that there is no significant correlation between symmetry and memorability scores. Figure 5 shows the scatter plots of memorability as a function of the composition related features and the sentiment related feature.

Furthermore, to address the question of relation between color and memorability, we transform the images from the RGB to the HSV color space. We calculate a 12-bin normalized histogram of the hue values in each image, where the position of the bin edges corresponds to the 30 degrees interval of 12 major colors in the HSV color wheel. Figure 6 shows the color wheel with Spearman's rank correlation coefficient for each of the 12 color subspaces. Although there is no significant correlation between hue and memorability, a weak positive correlation can be observed for warm colors, particularly orange and yellow.

In addition to addressing low level image features related to composition and color, we explore the relation of memorability to the visual sentiment evoked by the image. For this purpose, we use as feature extractor the Sentiment CNN model [27]. This CNN is a CaffeNet model fine-tuned for visual sentiment prediction on the DeepSent dataset [28], a set of 1269 Twitter images manually annotated as reflecting either positive or negative sentiment. We use the Sentiment CNN to extract the probability of negative and positive sentiment evoked by each of the 947 color field painting images. We find that for abstract images the probability of evoking positive sentiment has a moderately positive correlation with memorability. This result is particularly interesting because it indicates that specific visual properties of abstract images contribute not only to memorability, but also to the emotional response of the viewer.

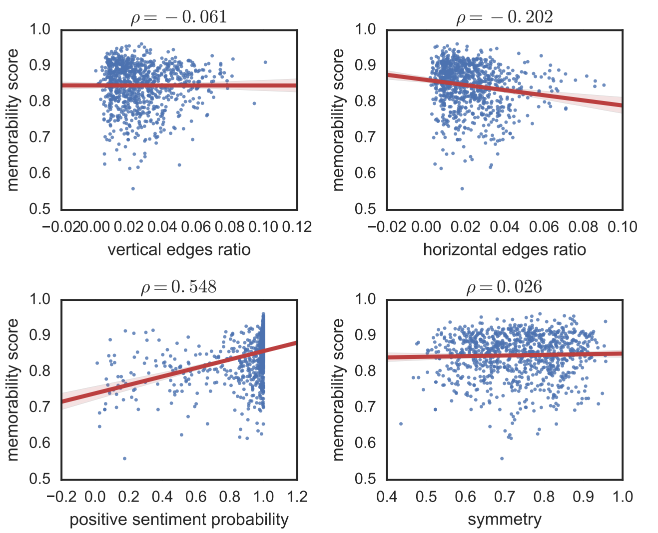

Fig. 5. Scatter plots show the correlation of memorability with (a) vertical edges ratio (b) horizontal edges ratio (c) positive sentiment probability (d) symmetry. In each scatter plot, each dot corresponds to one image. Red line is linear least squares fit.

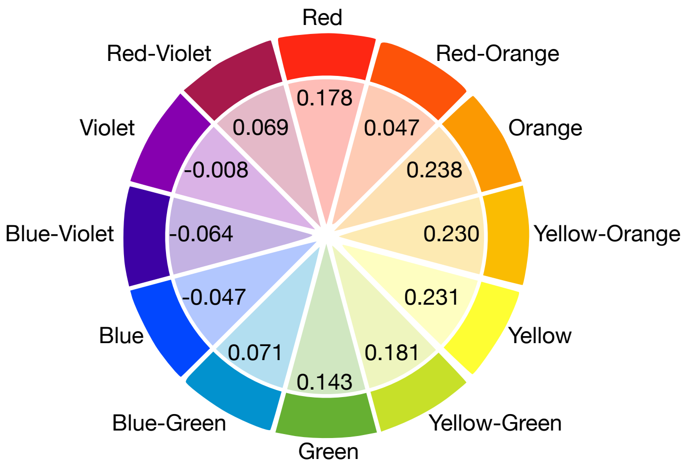

Fig. 6. Spearman's rank correlation between image memorability and each of the 12 color subspaces in the hue color wheel.

V. Conclusion

In this paper we address the question of memorability in fine art images through automated prediction of memorability scores using CNNs. We use the MemNet model, a CNN pretrained to predict image memorability of photographic images, in order to obtain predicted memorability scores of art images. Memorability is a complex feature and collecting memorability scores through human-driven experiments is time consuming and expensive for large-scale datasets. Because the MemNet model has near-human level of consistency in predicting memorability for variety of images, the results obtained for the art datasets can be considered relevant for interpreting the memorability of artworks. Regarding the artistic genre, we show that nude paintings and portraits are most memorable categories, while landscapes and marine paintings are the least memorable. In relation to artistic style, our experiments show that abstract art tends to be more memorable than figurative art. Furthermore, we provide an analysis of the correlation between memorability and image features related to composition, color and the visual sentiment response evoked by abstract images. In our future work we aim to further explore the question of relation within a larger set of features, with a particular focus on semantic and CNN-based image features.

978-1-5386-6979-2/18/$31.00 ©2018 European Union

Acknowledgment

This research has been partly supported by the European Regional Development Fund under the grant KK.01.1.1.01.0009 (DATACROSS).

References

[1] P. Isola, J. Xiao, A. Torralba, and A. Oliva, "What makes an image memorable?" in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 145–152.

[2] A. Khosla, A. S. Raju, A. Torralba, and A. Oliva, "Understanding and predicting image memorability at a large scale," in *Computer Vision (ICCV), 2015 IEEE International Conference on*. IEEE, 2015, pp. 2390–2398.

[3] N. van Noord, E. Hendriks, and E. Postma, "Toward discovery of the artist's style: Learning to recognize artists by their artworks," *IEEE Signal Processing Magazine*, vol. 32, no. 4, pp. 46–54, 2015.

[4] E. Cetinic and S. Grgic, "Automated painter recognition based on image feature extraction," in *ELMAR, 2013 55th International Symposium*. IEEE, 2013, pp. 19–22.

[5] T. E. Lombardi, *Classification of Style in Fine-art Painting*. Pace University, 2005.

[6] R. S. Arora and A. Elgammal, "Towards automated classification of fine-art painting style: A comparative study," in *Pattern Recognition (ICPR), 2012 21st International Conference on*. IEEE, 2012, pp. 3541–3544.

[7] S. Karayev, M. Trentacoste, H. Han, A. Agarwala, T. Darrell, A. Hertzmann, and H. Winnemoeller, "Recognizing image style," in *Proceedings of the British Machine Vision Conference*. BMVA Press, 2014.

[8] Z. Falomir, L. Museros, I. Sanz, and L. Gonzalez-Abril, "Categorizing paintings in art styles based on qualitative color descriptors, quantitative global features and machine learning (qart-learn)," *Expert Systems with Applications*, vol. 97, pp. 83–94, 2018.

[9] S. Agarwal, H. Karnick, N. Pant, and U. Patel, "Genre and style based painting classification," in *Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on*. IEEE, 2015, pp. 588–594.

[10] E. Cetinic and S. Grgic, "Genre classification of paintings," in *ELMAR, 2016 International Symposium*. IEEE, 2016, pp. 201–204.

[11] B. Saleh, A. Elgammal, J. Feldman, and A. Farhadi, "Toward a taxonomy and computational models of abnormalities in images," in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*. AAAI Press, 2016, pp. 3588–3596.

[12] E. J. Crowley and A. Zisserman, "In search of art." in *ECCV Workshops (1)*, 2014, pp. 54–70.

[13] A. Brachmann, E. Barth, and C. Redies, "Using cnn features to better understand what makes visual artworks special," *Frontiers in psychology*, vol. 8, p. 830, 2017.

[14] P. Isola, D. Parikh, A. Torralba, and A. Oliva, "Understanding the intrinsic memorability of images," in *Advances in Neural Information Processing Systems*, 2011, pp. 2429–2437.

[15] P. Isola, J. Xiao, D. Parikh, A. Torralba, and A. Oliva, "What makes a photograph memorable?" *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, no. 7, pp. 1469–1482, 2014.

[16] M. Mancas and O. Le Meur, "Memorability of natural scenes: The role of attention," in *Image Processing (ICIP), 2013 20th IEEE International Conference on*. IEEE, 2013, pp. 196–200.

[17] J. Kim, S. Yoon, and V. Pavlovic, "Relative spatial features for image memorability," in *Proceedings of the 21st ACM international conference on Multimedia*. ACM, 2013, pp. 761–764.

[18] P. Jing, Y. Su, L. Nie, and H. Gu, "Predicting image memorability through adaptive transfer learning from external sources," *IEEE Transactions on Multimedia*, vol. 19, no. 5, pp. 1050–1062, 2017.

[19] L. Goetschalckx, S. Vanmarcke, P. Moors, and J. Wagemans, "Are memorable images easier to categorize rapidly?" *Journal of Vision*, vol. 17, no. 10, pp. 98–98, 2017.

[20] X. Amengual, A. Bosch, and J. L. de la Rosa, "How to measure memorability and social interestingness of images: A review," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 31, no. 02, p. 1754004, 2017.

[21] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, "Learning deep features for scene recognition using places database," in *Advances in neural information processing systems*, 2014, pp. 487–495.

[22] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, pp. 675–678.

[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[24] J. Deng, A. Berg, S. Satheesh, H. Su, A. Khosla, and L. Fei-Fei, "Ilsvrc-2012, 2012," *URL http://www. image-net. org/challenges/LSVRC*, 2012.

[25] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.

[26] A. Brachmann and C. Redies, "Using convolutional neural network filters to measure left-right mirror symmetry in images," *Symmetry*, vol. 8, no. 12, p. 144, 2016.

[27] V. Campos, B. Jou, and X. Giro-i Nieto, "From pixels to sentiment: Fine-tuning cnns for visual sentiment prediction," *Image and Vision Computing*, vol. 65, pp. 15–22, 2017.

[28] Q. You, J. Luo, H. Jin, and J. Yang, "Robust image sentiment analysis using progressively trained and domain transferred deep networks." in *AAAI*, 2015, pp. 381–388.