Encodji: Encoding Gaze Data Into Emoji Space for an Amusing Scanpath Classification Approach ;)

Wolfgang Fuhl
Perception Engineering
University Tübingen
wolfgang.fuhl@uni-tuebingen.de

Efe Bozkir
Perception Engineering
University Tübingen
efe.bozkir@uni-tuebingen.de

Benedikt Hosp
Perception Engineering
University Tübingen
benedikt.hosp@uni-tuebingen.de

Nora Castner
Perception Engineering
University Tübingen
castnern@informatik.uni-tuebingen.de

David Geisler
Perception Engineering
University Tübingen
david.geisler@uni-tuebingen.de

Thiago C. Santini
Perception Engineering
University Tübingen
thiago.santini@uni-tuebingen.de

Enkelejda Kasneci
Perception Engineering
University Tübingen
enkelejda.kasneci@uni-tuebingen.de

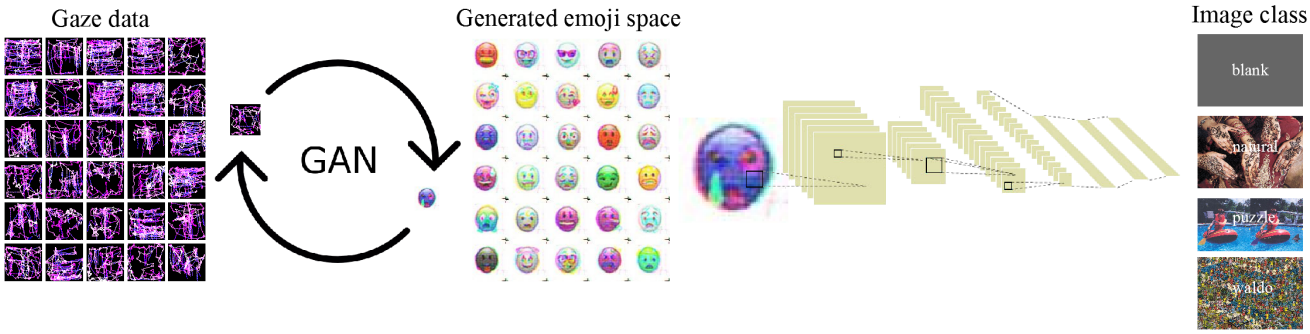

Figure 1: Our approach's workflow. The raw gaze data is encoded into Emoji space using a GAN generator, and the resulting Emoji is used for scanpath classification, improving classification accuracy.

ABSTRACT

To this day, a variety of information has been obtained from human eye movements, which holds an imense potential to understand and classify cognitive processes and states – e.g., through scanpath classification. In this work, we explore the task of scanpath classification through a combination of unsupervised feature learning and convolutional neural networks. As an amusement factor, we use an Emoji space representation as feature space. This representation is achieved by training generative adversarial networks (GANs) for unpaired scanpath-to-Emoji translation with a cyclic loss. The resulting Emojis are then used to train a convolutional neural network for stimulus prediciton, showing an accuracy improvement of more than five percentual points compared to the same network trained using solely the scanpath data. As a side effect, we also obtain novel unique Emojis representing each unique scanpath. Our goal is to demonstrate the applicability and potential of unsupervised feature learning to scanpath classification in a humorous and entertaining way.

ETRA '19, June 25–28, 2019, Denver , CO, USA

ACM ISBN 978-1-4503-6709-7/19/06…$15.00
https://doi.org/10.1145/3314111.3323074

CCS CONCEPTS

• **Computing methodologies** → **Batch learning**; **Neural networks**; *Reconstruction*;

KEYWORDS

Generative Adversarial Networks, Eye Tracking, Scanpath, Emoji, Image Generation

ACM Reference Format:
Wolfgang Fuhl, Efe Bozkir, Benedikt Hosp, Nora Castner, David Geisler, Thiago C. Santini, and Enkelejda Kasneci. 2019. Encodji: Encoding Gaze Data Into Emoji Space for an Amusing Scanpath Classification Approach ;). In *2019 Symposium on Eye Tracking Research and Applications (ETRA '19), June 25–28, 2019, Denver , CO, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3314111.3323074

1 INTRODUCTION

Eye movements can reveal information about a person's cognitive processes and states – e.g., identifying mental disorders, separating novices from experts, and detecting a performed task. Therefore, fixation and saccade spatio-temporal patterns, also called *scanpaths*,

hold valuable information. Multiple studies have identified such patterns in using gaze data: *a)* In the arts, viewing differences were found between experts and novices in abstract as well as realistic artworks [Zangemeister et al. 1995]. In addition, it was found that bottom-up features of the artwork and top-down beliefs affect the gaze behavior [Locher et al. 2015; Massaro et al. 2012]. *b)* In the medical domain, scanpath differences have been used to separate novices from expert microneurosurgeons [Eivazi et al. 2012; Kübler et al. 2015a] as well as radiologists [Manning et al. 2006; Van der Gijp et al. 2017]. A finer granular distinction was also made for dental students [Castner et al. 2018]. In addition, humans suffering from schizophrenia [Loughland et al. 2002] and autism spectrum disorder [Horley et al. 2004; Pelphrey et al. 2002] were also identified by their scanpath. Therefore, scanpaths have a high potential for training, diagnostic, and treatment monitoring in the medical domain. *c)* In the automotive domain, studies examining scanpath during driving were also made [Braunagel et al. 2017; Ji et al. 2004; Kasneci et al. 2014; Kübler et al. 2015b; Palinko et al. 2010]. Unsafe drivers can be robustly identified for humans suffering from visual field defects [Kasneci et al. 2014; Kübler et al. 2015b]. In autonomous driving, the take over readiness can also be estimated using the scanpath as an information source [Braunagel et al. 2017] as well as for measurements regarding cognitive load [Palinko et al. 2010] or fatigue [Ji et al. 2004] estimation.

2 RELATED WORKS

The first automated metric for scanpath classification was proposed in the nineties [Brandt and Stark 1997]. Following this publication, a variety of approaches have been presented and are summarized in [Anderson et al. 2015]. Modern approaches use machine learning in combination with statistically computed features [Boisvert and Bruce 2016; Crabb et al. 2014; French et al. 2017; Hoppe et al. 2018; Kübler et al. 2017; Liao et al. 2018; Zhang and Le Meur 2018]. Most algorithms rely on an agglomeration of time-aggregated features and sequence alignment. Newer methods use the transitions as an information source [Burch et al. 2018; Cristino et al. 2010; Dewhurst et al. 2018; Hoppe et al. 2018; Kübler et al. 2017]. These transitions have been shown to be a reliable feature for multiple tasks [Kübler et al. 2017] and were already applied in combination with Hidden Markov Models [Coutrot et al. 2018].

The recent progress in machine learning through convolutional neural networks (CNNs) [LeCun et al. 1998] has brought significant improvements in classification, detection, and segmentation. These improvements include the formulation of novel transposed convolution layers [Long et al. 2015], a softmax cross entropy loss formulation [Bishop 1994] as well as architectural enhancements. Modern architectures are based on residual [He et al. 2016], inception [Krizhevsky et al. 2012], or hybrid [Szegedy et al. 2016] layouts. These new architectures made it possible to train deeper and larger networks, which lead to high demand for annotated data. However, data annotation or developing realistic simulation able to generate labeled data are expensive and challenging to create, new ways to acquire training data are desirable. Thus, an alternative solution is to train CNNs to generate new training data out of existing images and easily generated data [Goodfellow et al. 2014] – i.e., generative adversarial networks (GANs). This training process required an image generator and a discriminator that decided whether the generated image is real or simulated [Goodfellow et al. 2014]. However, this training requires effort to determine the right learning parameters. With the cyclic loss function [Zhu et al. 2017], this effort has been significantly reduced and also makes unpaired training possible. In this work, we use GANs with the cyclic loss function for style transfer between images created based on gaze recordings and Emojis. Afterwards, the generated Emojis are used for scanpath classifiction using a CNN to predict the stimulus class.

3 METHOD

The proposed approach is illustrated in Figure 1. In this section, we describe the scanpath image generation, the employed networks architectures, and the training parameters and procedure. Each step is explained in a separate subsection for a better overview.

3.1 Scanpath Image Creation

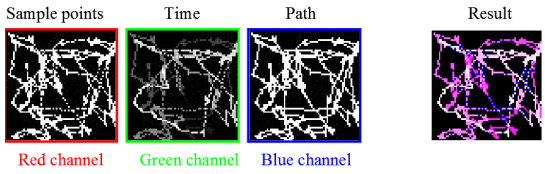

Figure 2: Scanpath image creation: the raw gaze data is encoded using RGB images, where the red, green, and blue channels hold the spatial, temporal, and connectivity path informations, respectively.

We employed the provided challenge data from the 2019 ACM Symposium on Eye Tracking Research & Applications [McCamy et al. 2014; Otero-Millan et al. 2008]. Only the free viewing data was used as the fixation data might unecessarily add extra noise due to errors in the eye movement classification algorithms. As shown in Figure 2, the raw data was encoded to an RGB image. The raw (normalized to the image size) gaze points coordinates were set as one in the red channel of the image (spatial information). In the green channel, these coodinates values were set based on the gaze point timestamps divided by the time of the recording (temporal information). In the blue channel, we interpolated all raw points to form connecting lines between subsequent gaze points (connectivity information).

3.2 Training Data Description

To generate the training sets, we used the images of the first four subjects (50% of the subjects from the provided dataset) created as described in section 3.1. These were augmented using elastic distortions [Simard et al. 2003] with scaling factors $\alpha = 44$ and $\sigma = 5$, resulting in 5,000 augmented scanpath images (henceforth the *Direct* training set). For the GANs, the *Direct* training set was used as source space. The target space was generated using the Emojis shown in Figure 3, augmented by a) randomly shuffling the color channel per Emoji and b) adding random blur and contrast changes. In total, we generated 5,000 augmented Emoji images.

Figure 3: Initial Emoji set for the training.

3.3 GANs Structure and Training

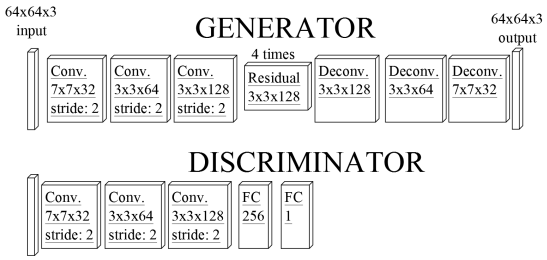

Figure 4: Generator and discriminator of the used GAN. Batch normalization (BN) and rectifier linear units (ReLu) are after each convolution and deconvolution block.

The used architecture for our GANs is shown in Figure 4. For simplification, we omitted the placement of rectifier linear units (ReLu) and batch normalization (BN), which are placed behind each convolution layer respectively. The training parameters for discriminators and generators were *BatchSize* = 1, *Solver* = *Adam*, *Momentum* = 0.5, *LearningRatePolicy* = *fixed*, and *LearninRateBase* = 0.0002. We trained for ninety epochs, which are 450,000 iterations using the Caffe framework [Jia et al. 2014]. The two generators and two discriminators were unsupervisedly trained using a cyclic loss function with *CycleLossWeight* = *10*. After the GANs were trained, the *Direct* training set was fed to the scanpath-to-Emoji generator network, resulting in 5,000 Emoji-encoded scanpath images (henceforth the *Emoji* training set). The best generated Emojis are shown in Figure 5; the largest variations are in the color space due to the training augmentation.

3.4 Classifier Structure and Training

We tried multiple classification architectures, and the best performing model (based on the *Direct* training set) is shown in Figure 6. This architecture was then trained to generate two classifiers: one using the *Direct* training set, and the other using the *Emoji* training set. Training was performed for ten epochs with the same hyperparameter set in both cases: *BatchSize* = 8, *Solver* = *Adam*, *Momentum* = 0.5, *LearningRatePolicy* = *fixed*, and *LearningRateBase* = 0.0001.

4 RESULTS

Table 1 shows the classification results. As can be seen, the Emoji generation can be useful to improve the classification accuracy. In

Figure 5: Best Emojis the GAN generated.

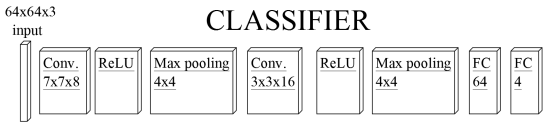

Figure 6: The architecture of our classification network.

our evaluation, we gained five percentual points over all classes (*Direct* 84% × *Emoji* 89%). The reason why this approach works and the general idea behind this improvement stems from unsupervised feature learning [Hu et al. 2018; Wu et al. 2018], in which every image gets a vector as target label. These vectors are equally spaced on a sphere to maximize the distance between each other. Afterwards, the vector can be classified linearly. In our case, this vector is represented by the Emoji space. Since our generated *Emoji* images do not follow any linearity, they cannot simply be separated linearly. Therefore, our generated images are more beautiful to look at.

Table 1: Results for the scanpath classifier using the *Direct* approach versus the *Emoji* one.

		Blank	Natural	Puzzle	Waldo	Accuracy
Direct	Blank	16	14	0	2	0.5
	Natural	2	50	0	4	0.89
	Puzzle	1	1	58	0	**0.96**
	Waldo	0	9	0	51	0.85
Emoji	Blank	19	12	0	1	**0.59**
	Natural	3	52	0	1	**0.92**
	Puzzle	1	1	58	0	**0.96**
	Waldo	0	2	1	57	**0.95**

5 CONCLUSION

In this work, we showed how to generate new emojis based on raw gaze data using GANs with a cyclic loss function, which lead to better scanpath classification results due to the maximization of the inter-feature-vector distance. This paper shows in a humorous way that modern machine learning approaches in combination with findings from the field of unsupervised learning can be used for scanpath classification. In addition, the approach can be used to generate new Emojis with amusing results (see Figure 5).

REFERENCES

Nicola C Anderson, Fraser Anderson, Alan Kingstone, and Walter F Bischof. 2015. A comparison of scanpath comparison methods. *Behavior Research Methods* 47, 4 (2015), 1377–1392.

Christopher M Bishop. 1994. *Mixture density networks.* Technical Report. Citeseer.

Jonathan FG Boisvert and Neil DB Bruce. 2016. Predicting task from eye movements: On the importance of spatial distribution, dynamics, and image features. *Neurocomputing* 207 (2016), 653–668.

Stephan A Brandt and Lawrence W Stark. 1997. Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cognitive Neuroscience* 9, 1 (1997), 27–38.

Christian Braunagel, David Geisler, Wolfgang Rosenstiel, and Enkelejda Kasneci. 2017. Online Recognition of Driver-Activity Based on Visual Scanpath Classification. *Intelligent Transportation Systems Magazine* 9, 4 (2017), 23–36.

Michael Burch, Kuno Kurzhals, Niklas Kleinhans, and Daniel Weiskopf. 2018. EyeMSA: exploring eye movement data with pairwise and multiple sequence alignment. In *Eye Tracking Research and Applications*. ACM, 52.

Nora Castner, Enkelejda Kasneci, Thomas Kübler, Katharina Scheiter, Juliane Richter, Thérése Eder, Fabian Hüttig, and Constanze Keutel. 2018. Scanpath comparison in medical image reading skills of dental students: distinguishing stages of expertise development. In *Eye Tracking Research and Applications*. ACM, 39.

Antoine Coutrot, Janet H Hsiao, and Antoni B Chan. 2018. Scanpath modeling and classification with hidden Markov models. *Behavior Research Methods* 50, 1 (2018), 362–379.

David P Crabb, Nicholas D Smith, and Haogang Zhu. 2014. What's on TV? Detecting age-related neurodegenerative eye disease using eye movement scanpaths. *Frontiers in Aging Neuroscience* 6 (2014), 312.

Filipe Cristino, Sebastiaan Mathôt, Jan Theeuwes, and Iain D Gilchrist. 2010. ScanMatch: A novel method for comparing fixation sequences. *Behavior Research Methods* 42, 3 (2010), 692–700.

Richard Dewhurst, Tom Foulsham, Halszka Jarodzka, Roger Johansson, Kenneth Holmqvist, and Marcus Nyström. 2018. How task demands influence scanpath similarity in a sequential number-search task. *Vision Research* 149 (2018), 9–23.

Shahram Eivazi, Roman Bednarik, Markku Tukiainen, Mikael von und zu Fraunberg, Ville Leinonen, and Juha E Jääskeläinen. 2012. Gaze behaviour of expert and novice microneurosurgeons differs during observations of tumor removal recordings. In *Eye Tracking Research and Applications*. ACM, 377–380.

Robert M French, Yannick Glady, and Jean-Pierre Thibaut. 2017. An evaluation of scanpath-comparison and machine-learning classification algorithms used to study the dynamics of analogy making. *Behavior Research Methods* 49, 4 (2017), 1291–1302.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in Neural Information Processing Systems*. MIT Press, 2672–2680.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Computer Vision and Pattern Recognition*. IEEE, 770–778.

Sabrina Hoppe, Tobias Loetscher, Stephanie A Morey, and Andreas Bulling. 2018. Eye movements during everyday behavior predict personality traits. *Frontiers in Human Neuroscience* 12 (2018), 105.

Kaye Horley, Leanne M Williams, Craig Gonsalvez, and Evian Gordon. 2004. Face to face: visual scanpath evidence for abnormal processing of facial expressions in social phobia. *Psychiatry Research* 127, 1-2 (2004), 43–53.

Lanqing Hu, Meina Kan, Shiguang Shan, and Xilin Chen. 2018. Duplex generative adversarial network for unsupervised domain adaptation. In *Computer Vision and Pattern Recognition*. IEEE, 1498–1507.

Qiang Ji, Zhiwei Zhu, and Peilin Lan. 2004. Real-time nonintrusive monitoring and prediction of driver fatigue. *Transactions on Vehicular Technology* 53, 4 (2004), 1052–1068.

Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. 2014. Caffe: Convolutional Architecture for Fast Feature Embedding. *CoRR* abs/1408.5093 (2014). arXiv:1408.5093 http://arxiv.org/abs/1408.5093

Enkelejda Kasneci, Katrin Sippel, Kathrin Aehling, Martin Heister, Wolfgang Rosenstiel, Ulrich Schiefer, and Elena Papageorgiou. 2014. Driving with binocular visual field loss? A study on a supervised on-road parcours with simultaneous eye and head tracking. *PLOS ONE* 9, 2 (2014), e87470.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*. MIT Press, 1097–1105.

Thomas Kübler, Shahram Eivazi, and Enkelejda Kasneci. 2015a. Automated visual scanpath analysis reveals the expertise level of micro-neurosurgeons. In *Medical Image Computing and Computer Assisted Intervention Workshop on Interventional Microscopy*. Springer, 1–8.

Thomas C Kübler, Enkelejda Kasneci, Wolfgang Rosenstiel, Martin Heister, Kathrin Aehling, Katja Nagel, Ulrich Schiefer, and Elena Papageorgiou. 2015b. Driving with glaucoma: task performance and gaze movements. *Optometry and Vision Science* 92, 11 (2015), 1037–1046.

Thomas C Kübler, Colleen Rothe, Ulrich Schiefer, Wolfgang Rosenstiel, and Enkelejda Kasneci. 2017. SubsMatch 2.0: Scanpath comparison and classification based on subsequence frequencies. *Behavior Research Methods* 49, 3 (2017), 1048–1064.

Yann LeCun, Léon Bottou, Yoshua Bengio, Patrick Haffner, et al. 1998. Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 11 (1998), 2278–2324.

Hua Liao, Weihua Dong, Haosheng Huang, Georg Gartner, and Huiping Liu. 2018. Inferring user tasks in pedestrian navigation from eye movement data in real-world environments. *International Journal of Geographical Information Science* 33, 4 (2018), 1–25.

Paul Locher, Elizabeth Krupinski, and Alexandra Schaefer. 2015. Art and authenticity: Behavioral and eye-movement analyses. *Psychology of Aesthetics, Creativity, and the Arts* 9, 4 (2015), 356.

Jonathan Long, Evan Shelhamer, and Trevor Darrell. 2015. Fully convolutional networks for semantic segmentation. In *Computer Vision and Pattern Recognition*. IEEE, 3431–3440.

Carmel M Loughland, Leanne M Williams, and Evian Gordon. 2002. Visual scanpaths to positive and negative facial emotions in an outpatient schizophrenia sample. *Schizophrenia Research* 55, 1-2 (2002), 159–170.

David Manning, Susan Ethell, Tim Donovan, and Trevor Crawford. 2006. How do radiologists do it? The influence of experience and training on searching for chest nodules. *Radiography* 12, 2 (2006), 134–142.

Davide Massaro, Federica Savazzi, Cinzia Di Dio, David Freedberg, Vittorio Gallese, Gabriella Gilli, and Antonella Marchetti. 2012. When art moves the eyes: a behavioral and eye-tracking study. *PLOS ONE* 7, 5 (2012), e37285.

Michael B McCamy, Jorge Otero-Millan, Leandro Luigi Di Stasi, Stephen L Macknik, and Susana Martinez-Conde. 2014. Highly informative natural scene regions increase microsaccade production during visual scanning. *Journal of Neuroscience* 34, 8 (2014), 2956–2966.

Jorge Otero-Millan, Xoana G Troncoso, Stephen L Macknik, Ignacio Serrano-Pedraza, and Susana Martinez-Conde. 2008. Saccades and microsaccades during visual fixation, exploration, and search: foundations for a common saccadic generator. *Journal of Vision* 8, 14 (2008), 21–21.

Oskar Palinko, Andrew L Kun, Alexander Shyrokov, and Peter Heeman. 2010. Estimating cognitive load using remote eye tracking in a driving simulator. In *Eye Tracking Research and Applications*. ACM, 141–144.

Kevin A Pelphrey, Noah J Sasson, J Steven Reznick, Gregory Paul, Barbara D Goldman, and Joseph Piven. 2002. Visual scanning of faces in autism. *Journal of Autism and Developmental Disorders* 32, 4 (2002), 249–261.

Ashish Shrivastava, Tomas Pfister, Oncel Tuzel, Joshua Susskind, Wenda Wang, and Russell Webb. 2017. Learning from simulated and unsupervised images through adversarial training. In *Computer Vision and Pattern Recognition*. IEEE, 2107–2116.

Patrice Y Simard, David Steinkraus, John C Platt, et al. 2003. Best practices for convolutional neural networks applied to visual document analysis.. In *International Conference on Document Analysis and Recognition*, Vol. 3. IEEE.

Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. 2016. Rethinking the inception architecture for computer vision. In *Computer Vision and Pattern Recognition*. IEEE, 2818–2826.

A Van der Gijp, CJ Ravesloot, H Jarodzka, MF Van der Schaaf, IC Van der Schaaf, Jan PJ van Schaik, and Th J Ten Cate. 2017. How visual search relates to visual diagnostic performance: a narrative systematic review of eye-tracking research in radiology. *Advances in Health Sciences Education* 22, 3 (2017), 765–787.

Zhirong Wu, Yuanjun Xiong, Stella X Yu, and Dahua Lin. 2018. Unsupervised feature learning via non-parametric instance discrimination. In *Computer Vision and Pattern Recognition*. IEEE, 3733–3742.

WH Zangemeister, Keith Sherman, and Lawrence Stark. 1995. Evidence for a global scanpath strategy in viewing abstract compared with realistic images. *Neuropsychologia* 33, 8 (1995), 1009–1025.

A Tianyi Zhang and B Olivier Le Meur. 2018. How Old Do You Look? Inferring Your Age from Your Gaze. In *International Conference on Image Processing*. IEEE, 2660–2664.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *International Conference on Computer Vision*. IEEE, 2223–2232.