



HAL
open science

Eyes on Emotions: What Makes Faces So Captivating?

Quentin Sénant, Aurore Philippe, Lea Entzmann, Nathalie Guyader, Martial Mermillod

► **To cite this version:**

Quentin Sénant, Aurore Philippe, Lea Entzmann, Nathalie Guyader, Martial Mermillod. Eyes on Emotions: What Makes Faces So Captivating?. ECVF 2024 - 46th European Conference on Visual Perception, Aug 2024, Aberdeen, United Kingdom. . <hal-04689081>

HAL Id: hal-04689081

<https://hal.science/hal-04689081v1>

Submitted on 5 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Eyes on Emotions: What Makes Faces So Captivating?

Quentin Sénant⁽¹⁾, Aurore Philippe⁽¹⁾, Léa Entzmann⁽³⁾, Nathalie Guyader⁽²⁾, Martial Mermillod⁽¹⁾

(1) Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LPNC, 38000 Grenoble, France. (2) Univ. Grenoble Alpes, CNRS, GIPSA-lab, 38000 Grenoble, France. (3) Icelandic Vision Lab, School of Health Sciences, University of Iceland, Reykjavík, Iceland



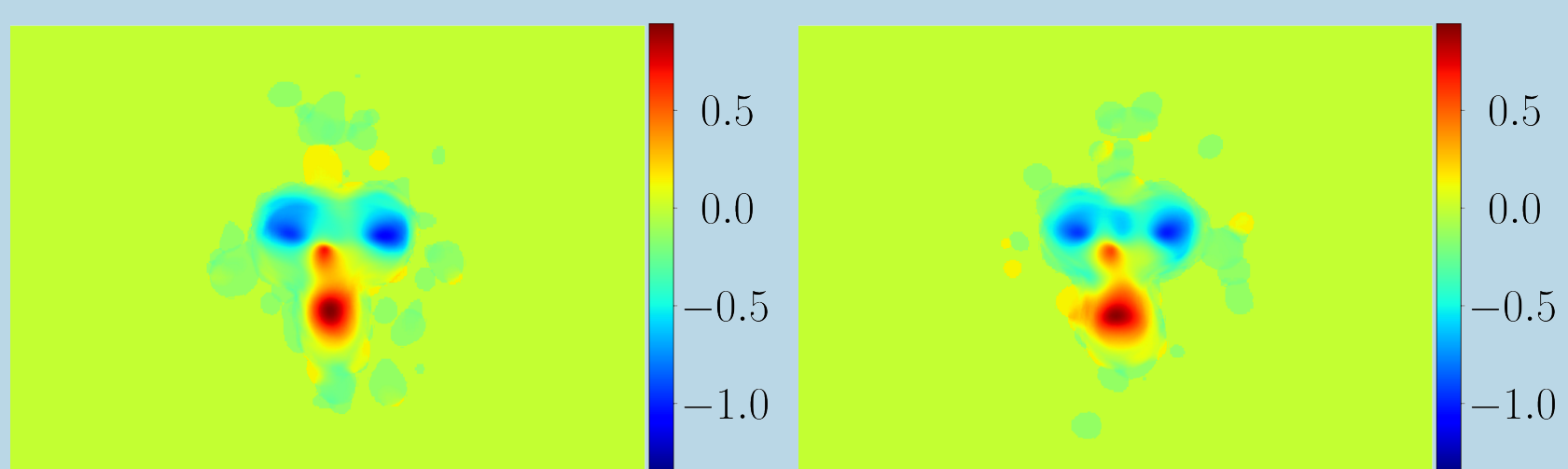
Introduction

Attention maps can be used to predict the regions most likely to attract the eye, thus highlighting the so-called “salient regions”. These regions depend both on the statistical properties of the scenes and the task of the observers. Several bottom-up models have been proposed to predict salient regions and more recently, **DeepGaze IIE**, a deep neural network model, has been trained directly onto human eye-tracking data to generalize them to unseen images but ignores task effects on eye fixations. Other methods, such as LayerCAM [4] provide maps that enhance the regions used by CNNs for a specific task.

In this context, we are interested in studying the salient regions of emotional faces. Emotional faces are specific stimuli that strongly attract gaze [1], especially on the eye and mouth. But it is unclear whether humans focus on these regions due to the performed task of expression recognition or due to the fact that these are “salient”. In a previous work [2], we demonstrated that CNNs maps highlight these and predict eye behaviour. Here, we:

- develop saliency maps that incorporates both bottom-up and top-down information.
- compare visual ROIs between eye-tracking and models’ data for emotional faces.
- evaluate CNN maps for predicting visual fixations during two tasks.

Cohen’s maps Emotion – Gender for Angry and Happy faces



Hypotheses

- **Eye fixations** focus on emotion- or gender-diagnostic information based on the categorization task.
- **Incorporating** task-specific information into saliency models improves prediction of eye fixation.
- **CNNs**, with task information, outperforms state-of-the-art saliency models.

References

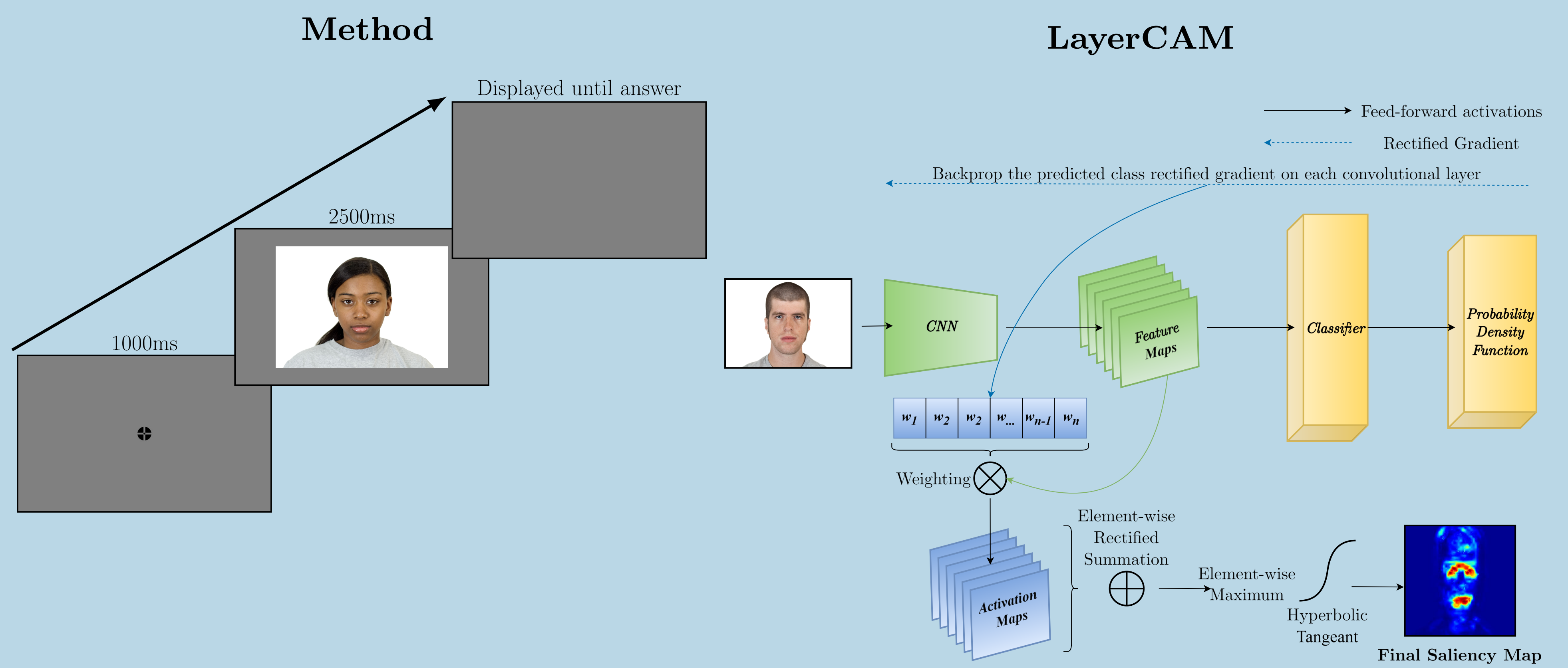
- [1] Schurgin, M. W., Nelson, J. S., Iida, S., Ohira, H., Chiao, J. Y., & Franconeri, S. L. (2014). Eye movements during emotion recognition in faces. *Journal of Vision*, 14 13, 14.
- [2] Entzmann, L., Guyader, N., Kauffmann, L., Peyrin, C., & Mermillod, M. (2021). Behavioral and computational assessment of the role of spatial frequencies in emotion detection. *ECVP 2021 - European Conference on Visual Perception*.
- [3] VanRullen, R. (2003). Visual saliency and spike timing in the ventral visual pathway. *Journal of Physiology-Paris*, 97, 365–377.
- [4] Jiang, P.-T., Zhang, C.-B., Hou, Q., Cheng, M.-M., & Wei, Y. (2021). LayerCAM: Exploring Hierarchical Class Activation Maps for Localization. *IEEE Transactions on Image Processing*, 30, 5875–5888.

Acknowledgements

This research was supported by the GATES project (Grenoble ATtractiveness and ExcellenceS), funded by the French government’s Programme d’Investissement Avenir and implemented by ‘ANR France 2030’ (ANR-22-EXES-0001) and was partially supported by MIAI @ Grenoble Alpes, (ANR-19-P3IA-0003).

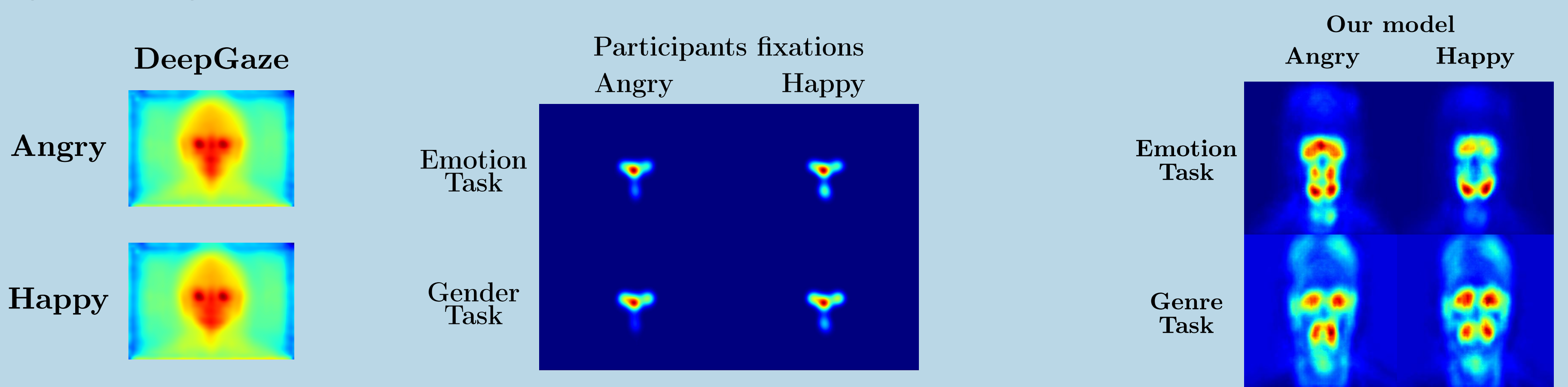
Method

We submitted gender and emotion categorization tasks to thirty-three participants and collected their eye-tracking data. Half of them saw fearful and neutral faces, half of them saw angry and happy faces. Each group performed the two tasks on the same set of stimuli and the order was counterbalanced between participants. We optimized two convolutional neural networks for the realization of the same tasks as those performed by participants. After the training phase, we collected LayerCAM [4] class activation maps. Then, we compared the tasks’ effects on eye fixation distribution and the performance of models’ saliency maps.

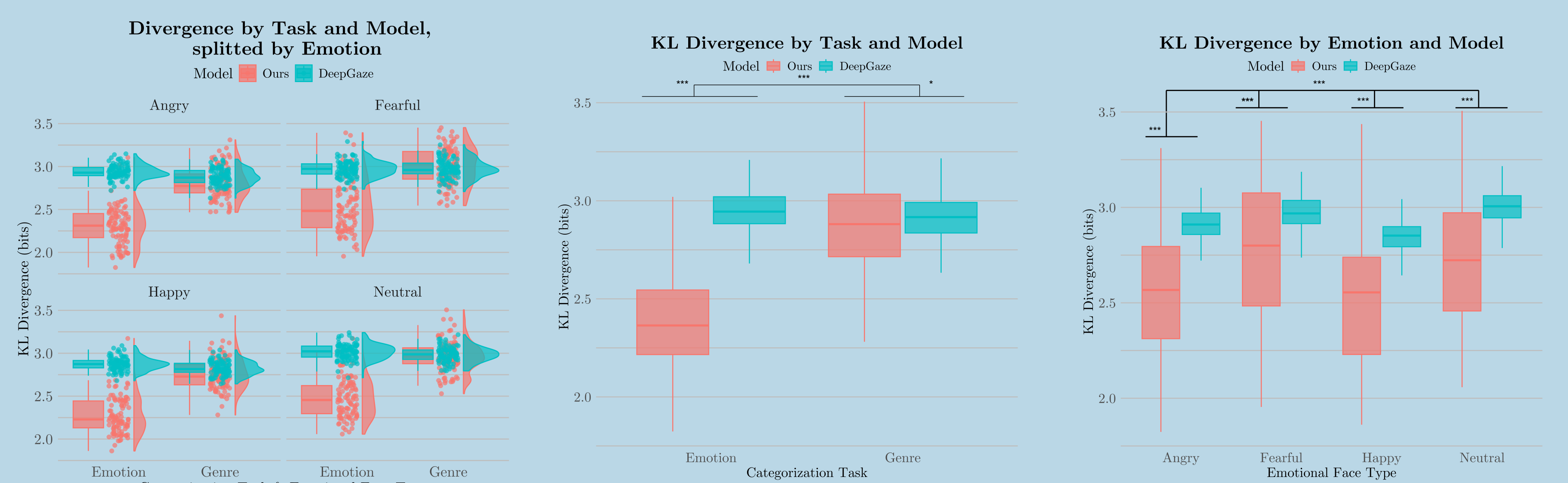


Results

There is an **effect of the task (Gender vs. Emotion categorization)** on the accuracies of participants such that they are, on average, more accurate when performing gender recognition than emotion recognition. This effect **interacts with the type of participant group**, as there is no significant effect of the task on the scores for the happy and angry faces group, but this effect is strong for the group that saw fearful and neutral faces.



The DeepGaze model, as compared to ours, does not take into account the task effects. Therefore, we tested whether it is a better model for the eye fixation densities for emotion categorization or gender categorization. Our results demonstrate that taking into account the task allows for better saliency map predictions. In particular, when analyzing Kullback-Leibler divergences (*lower is better*), we found both a **task by model** and an **emotion by model** interaction.



Conclusion

We demonstrate that:

- **Eye movements** are task-dependant, even in the very first few saccades.
- **Incorporating** top-down information into visual attention maps is needed to correctly predict eye movements.
- **CNNs** constitute good candidates to build such attention maps.