

Étude comparative de méthodes de reconnaissance des émotions à partir d’expressions faciales

Rim El Cheikh*, H el ene Tran*,** Issam Falih*, Engelbert Mephu Nguifo*

* Universit e Clermont-Auvergne, CNRS, Mines de Saint- tienne,
Clermont Auvergne INP, LIMOS, 63000 Clermont-Ferrand, France

{rim.el_cheikh, helene.tran}@doctorant.uca.fr

{issam.falih, engelbert.mephu_nguifo}@uca.fr

** Jeolis Solutions, 63000 Clermont-Ferrand, France

1 Introduction

La recherche en reconnaissance automatique des  motions est active depuis de nombreuses d ecennies et ses applications sont diverses, telles que la sant e et le divertissement. Notre  tude porte sur les m ethodes qui pr edisent les  motions   partir d’images d’expressions faciales (FER). Trois mod eles d’ tat de l’art pour les t aches FER ont  t e s electionn es pour  tre exp eriment es. Ils divergent dans leurs architectures et la m ethodologie utilis ee pour am eliorer la qualit e de l’inf erence des  motions. Nos exp eriences fournissent une comparaison  quitable de leurs performances sur trois ensembles de donn ees qui diff erent en termes de taille, de m ethodologie de collecte d’images et de distribution des classes.

2 Exp erimentations et discussion

Les r eseaux de neurones  tudi es sont ESR (Siqueira et al., 2020), SCN (Wang et al., 2020) et DACL (Farzaneh et Qi, 2021), qui utilisent respectivement l’apprentissage par ensembles, la relabellisation des donn ees incertaines, et l’attention. Ils sont  valu es par validation crois ee 5-fold sur les trois jeux de donn ees FER+¹, AffectNet² et CK+³. Une description d etaill ee des r eseaux, des donn ees et du protocole exp erimental se trouve dans notre article (El Cheikh et al., 2022).

Le tableau 1 reporte les mesures de performance des mod eles sur chaque jeu de donn ees. DACL fournit les meilleurs scores sur FER+ et AffectNet, deux jeux de donn ees contenant des expressions faciales spontan ees. Cela sugg ere que son module d’attention aide   se focaliser sur les r egions pertinentes pour d eduire l’ motion. ESR donne les meilleures performances sur CK+ qui contient des  motions act ees. Le r eseau utilise des repr esentations partag ees pour l’entranement de chaque branche (Siqueira et al., 2020), ce qui s’av ere efficace dans un contexte o  les  motions sont pos ees et intentionnelles.

1. <https://github.com/microsoft/FERPlus/>

2. <http://mohammadmahoor.com/affectnet/>

3. <https://www.jeffcohn.net/resources/>

Comparaison de méthodes de reconnaissance d'émotions faciales

		TJ	TJP	PR	RP	F1	AUC ROC
FER+	ESR	0.857	0.617	0.855	0.857	0.856	0.937
	SCN	0.810	0.520	0.808	0.810	0.809	0.956
	DACL	0.867	0.647	0.863	0.867	0.865	0.973
AffectNet	ESR	0.648	0.439	0.626	0.648	0.637	0.821
	SCN	0.651	0.390	0.622	0.651	0.636	0.894
	DACL	0.664	0.429	0.633	0.664	0.648	0.901
CK+	ESR	0.915	0.888	0.922	0.915	0.918	0.945
	SCN	0.820	0.703	0.798	0.820	0.808	0.962
	DACL	0.846	0.790	0.843	0.846	0.844	0.951

TAB. 1 – Moyennes des métriques de performance (taux de justesse (TJ), TJ pondéré (TJP), précision (PR), rappel (RP), score F1, score AUC ROC) sur les données de test.

Les matrices de confusion présentées dans notre article (El Cheikh et al., 2022) mettent en évidence la difficulté des modèles à différencier certaines émotions. Par exemple, de nombreux échantillons sont classés "neutre" alors qu'ils contiennent une émotion, notamment lorsque celles-ci sont naturelles (FER+ et AffectNet). En particulier, l'entraînement sur FER+ induit beaucoup d'erreurs en prédisant "neutre" pour les données étiquetées "triste", mais pas pour ceux étiquetés "surprise", bien que "triste" et "surprise" soient en proportions très proches. Cela suggère que l'expression neutre et la tristesse ont des caractéristiques physiques semblables.

3 Conclusion

Cette étude comparative montre l'impact de l'architecture neuronale sur la classification des émotions, actées ou spontanées. Des défis ont été identifiés, tels que la sous-représentation de certaines classes d'émotion et l'identification ambiguë de l'émotion dans un contexte non contrôlé. Étendre les expériences à plus de modèles et d'ensembles de données fournirait une référence fiable pour choisir un modèle FER adapté en fonction de l'application souhaitée.

Références

- El Cheikh, R., H. Tran, I. Falih, et E. M. Nguifo (2022). A comparative study of emotion recognition methods using facial expressions. <https://arxiv.org/abs/2212.03102>.
- Farzaneh, A. H. et X. Qi (2021). Facial expression recognition in the wild via deep attentive center loss. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pp. 2402–2411.
- Siqueira, H., S. Magg, et S. Wermter (2020). Efficient facial feature learning with wide ensemble-based convolutional neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, Volume 34, pp. 5800–5809.
- Wang, K., X. Peng, J. Yang, S. Lu, et Y. Qiao (2020). Suppressing uncertainties for large-scale facial expression recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 6897–6906.