A Two-Step Learning Method for Detecting Landmarks on Faces from Different Domains

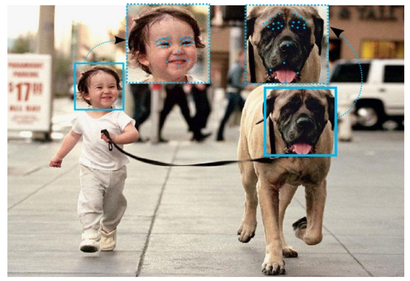
Bruna V. Frade Erickson R. Nascimento
{brunafrade, erickson}@dcc.ufmg.br
Universidade Federal de Minas Gerais (UFMG), Brazil

Sum-up

- The animation in the non-human face needs of the landmarks detection.
- The methods of the landmarks detection use machine learning techniques and require many labeled annotation image.
- Many labeled data for a human face and little or any labeled data for the non-human face.
- We propose a method (2Steps) for landmarks detection in domains different face using an approach that encodes features for the multi-task learning: i) source output landmarks (human). ii) reconstruction target face (non-human).
- Learning objective of our method is related features of the two domains labeled and non-labeled for decrease the labeled dataset dependency for landmarks detection in the non-human face.

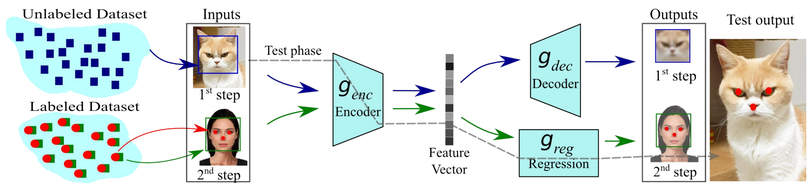
Landmarks definition

➢Landmarks are discriminative locations in the image. In faces a landmarks locate regions comprising the eyes,eyebrows, mouth and nose.



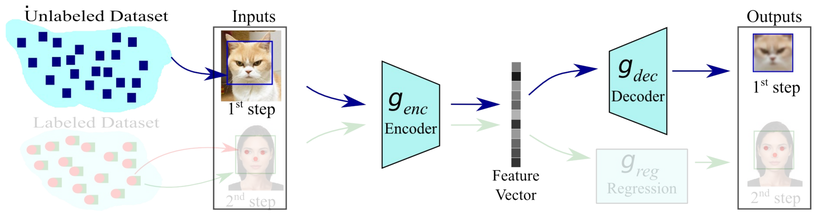
The 2Steps learning approach

➢Our formulation is based on a two-step learning approach, where in first step it learn to reconstruct images from D^{target} domain using an unsupervised strategy and second step it solve regression problem in a supervised way predicting the landmarks coordinates in D^{source} domain..



First step

➢The face features are encoded using $g_{enc}(u_i)$ function and the decoded with $g_{dec}(g_{enc}(u_i))$

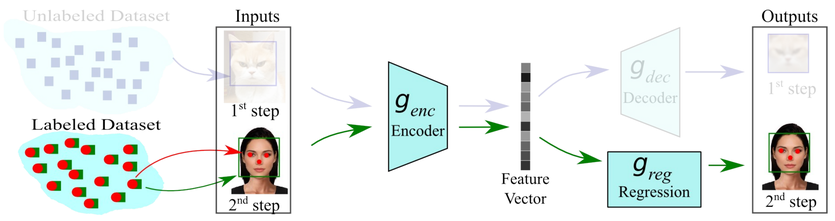


Let:

$$\mathcal{L}_{rec} = \frac{1}{n}\sum_{i=1}^{n} \|\mathbf{u}_i - g_{dec}(g_{enc}(\mathbf{u}_i))\|^2$$

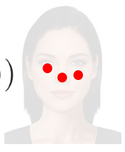
Second step

➢The face features are extracted using $g_{enc}(u_i)$ function and the landmarks are detected using $g_{reg}(g_{enc}(x_i))$ regression function.



Let:

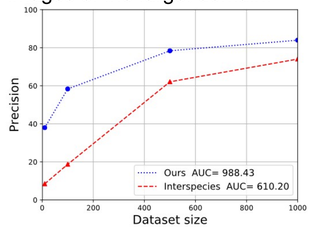
$$\mathcal{L}_{reg} = \frac{1}{m}\sum_{i=1}^{m} \mathrm{MAE}(\mathbf{y}_i - g_{reg}(g_{enc}(\mathbf{x}_i)))$$



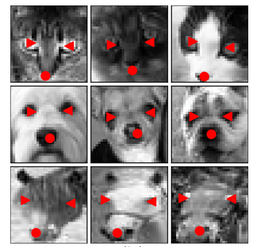
Ground truth Prediction

Results

➢Interspecies [1]: Source human, target horse. (a) Test data precision by varying the dataset size in the train. (b) Landmarks prediction using model trained with 100 images of the target domain.

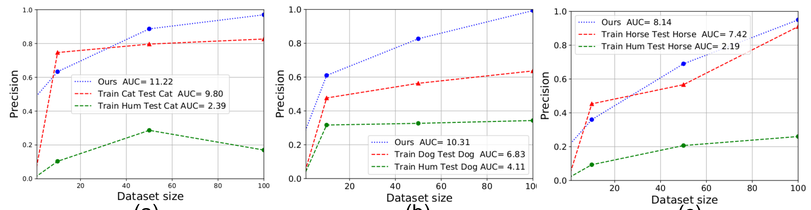


(a)



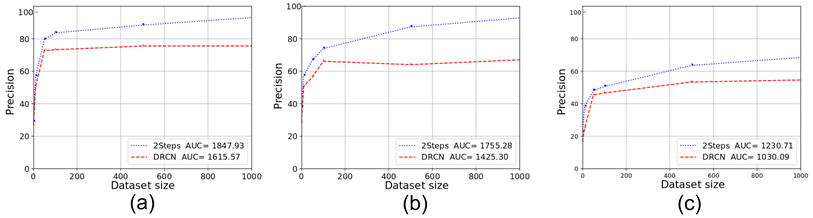
(b)

➢ConvNet: Ours supervised step: Varying dataset size in the train for the target domain (a) cat, (b) dog, (c) horse in the blue curve and human faces for the source domain. The red curve using only the target domain. Green curve using only human faces.



(a) (b) (c)

➢DRCN [3]: Varying dataset size in the train for the domain target (a) cat, (b) dog, (c) horse in the blue curve and human faces for the source domain.



(a) (b) (c)

References

[1] Maheen Rashid, Xiuye Gu, and Yong Jae Lee, “Interspecies knowledge transfer for facial keypoint detection,” Computer Vision and Pattern Recognition(CVPR), 2017.
[3] Muhammad Ghifary, W Bastiaan Kleijn, Mengjie Zhang, David Balduzzi, and Wen Li, “Deep reconstruction classification networks for unsupervised domain adaptation,” in European Conference on Computer Vision. Springer, 2016, pp. 597–613.