A Supplementary Material to the Paper: Deep Covariance Descriptors for Facial Expression Recognition

Naima Otberdout[1]
naima.otberdout@um5s.net.ma

Anis Kacem[2]
anis.kacem@imt-lille-douai.fr

Mohamed Daoudi[2]
mohamed.daoudi@imt-lille-douai.fr

Lahoucine Ballihi[1]
lahoucine.ballihi@um5.ac.ma

Stefano Berretti[3]
stefano.berretti@unifi.it

[1] LRIT - CNRST URAC 29
Rabat IT Center,
Mohammed V University in Rabat
Faculty of sciences, Rabat, Morocco

[2] IMT Lille-Douai, Univ. Lille,
CNRS, UMR 9189 CRIStAL,
Lille, France

[3] Media Int. and Com. Center (MICC),
University of Florence,
Florence, Italy

In this supplementary material, we present further details on the conducted experiments. In particular, we provide visualizations of:

1 Global DCNN features and their covariance descriptors

Figure 1 shows four selected feature maps (chosen from 512 FMs) extracted with the ExpNet model for two subjects of the Oulu-CASIA dataset (happy and surprise expressions). We also show the global covariance descriptor relative to the 512 feature maps as a 2D image. Common patterns can be observed in the covariance descriptors computed for similar expressions, *e.g.*, the dominant colors in the covariance descriptors of happy expression (left panel) are pink/purple, while being blue in the covariance descriptors of surprise expression (right panel).

2 Local DCNN features and their covariance descriptors

Figure 2 shows the four local regions detected on the input facial image on the left; then, landmarks and regions are shown on four selected feature maps, as mentioned in Section 3.2 of the paper. These FMs are selected from 512. The covariance descriptors relative to each detected region are shown in Figure 3. We can observe that each local covariance descriptor captures different patterns.
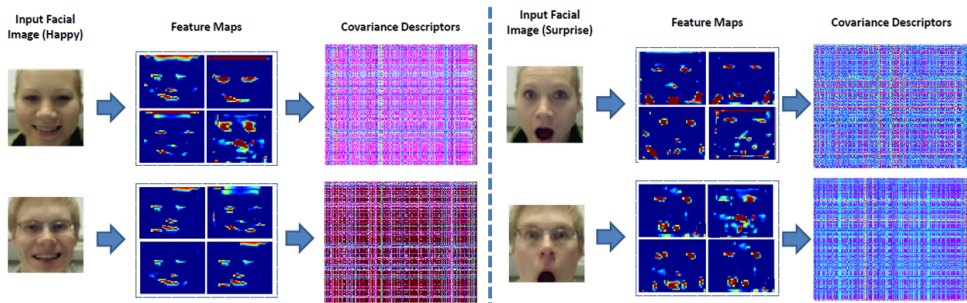
Figure 1: Visualization of some feature maps (ExpNet) and their corresponding covariance descriptors for two subjects from the Oulu-CASIA dataset conveying happy and surprise expressions. We show four feature maps (chosen from 512 feature maps) for each example image. The corresponding covariance descriptors are computed over the 512 FMs. Best seen in color.
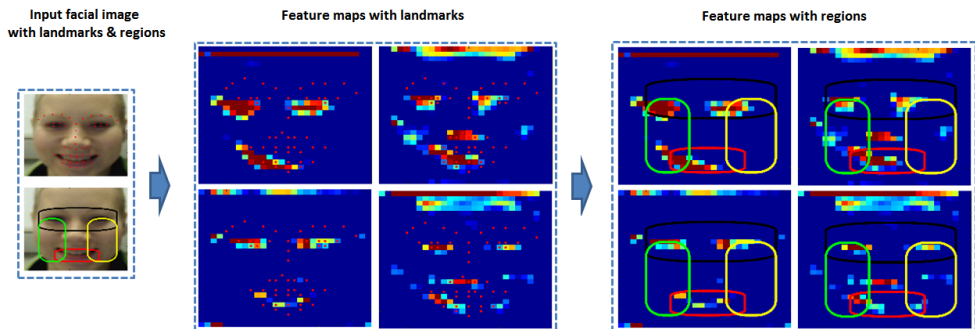

Figure 2: Visualization of the detected facial landmarks and regions on the input facial image and mapped on four selected feature maps (from 512). Best seen in color.

3 Failure cases of facial landmark detection on SFEW dataset

Figure 4 exhibits some failure and success cases of facial landmark and region detection on the input facial images. In the left panel of this figure, we show examples from the Oulu-CASIA and SFEW datasets, where the landmark and region detection succeeded. In the right panel, we show four failure examples for landmark and region detection in the SFEW dataset. We noticed that this step failed on $\sim 30\%$ of the facial images of SFEW. This explains why we do not obtain improvements by combining local and global covariance descriptors on this dataset.
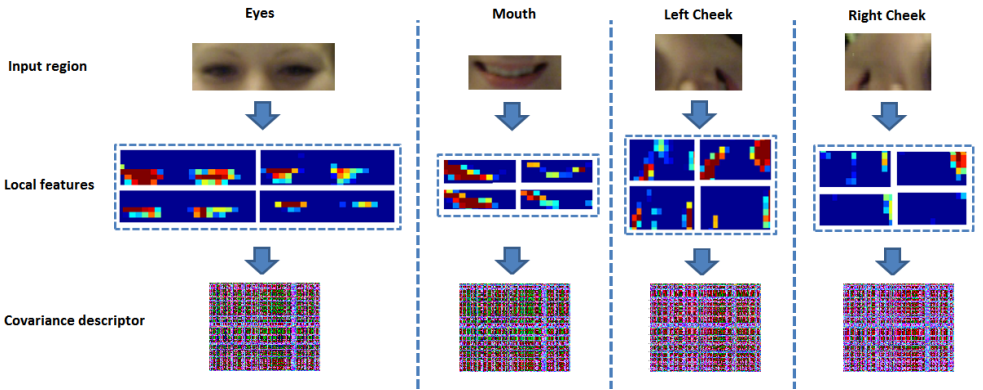
Figure 3: Visualization of the extracted regions on four feature maps and their corresponding covariance descriptors. Best seen in color.
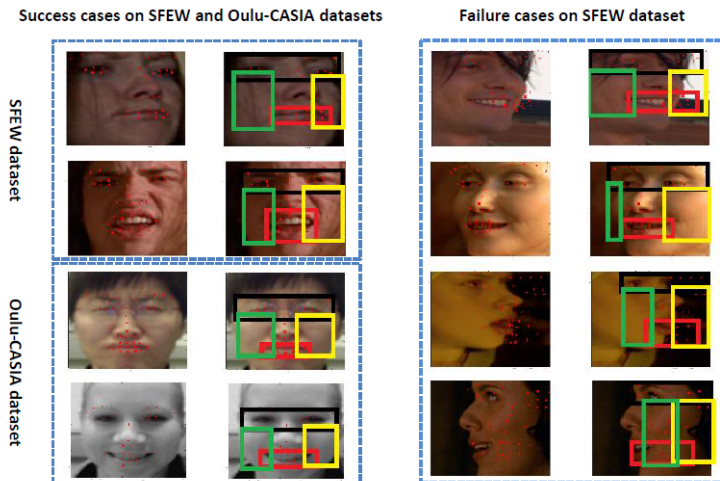

Figure 4: Examples of facial landmark and region detection on the SFEW and Oulu-CASIA datasets, with some failure cases for the SFEW dataset. For each example, the image on the left shows the aligned face with its landmark points, while the image on the right represents the aligned face with its detected regions.