Deep Region and Multi-label Learning for Facial Action Unit Detection

Kaili Zhao¹ Wen-Sheng Chu² Honggang Zhang¹
¹Beijing University of Posts and Telecommunications
²Carnegie Mellon University

Alternative Methods

▲ Alternative CNNs
- AlexNet [1]: An AlexNet trained with multi-label loss.
- Locally Connected Network (LCN) [3]: ConvNet architecture with conv6-conv7 replaced with locally connected layers.
- Regular ConvNet: DRML without the region layer.
- DRML: The proposed architecture.

▲ Comparisons
- Methods: End-to-end trainable, multi-label learning, Learning representations.
- Performance: Non-linearity update, Online.
- Techniques: APL, LCN, APL, [3], [3], [3].

Regions v.s. Patches

▲ Importance between regions and patches
- Active region: Magnitude of per-pixel gradients with respect to a particular AU, ie, saliency map [8].
- Active patch: L2 norm for pair-AU classifier, ie, [3] and APL [4].

▲ Learned regions and patches
- The face images were selected manually from the CK+ dataset [8].

Experiments

▲ Settings
- Prediction: 6D feature + Linear SVM.
- Metrics: P1 score and AUC.

▲ Scenarios
- Within-dataset: 5-fold partition for training/valid sets.
- Cross-dataset: 7-fold partition.

▲ Convergence of DRML
- Convergence curves for DRML and AlexNet for Facial Action Unit Detection.

▲ BIP4D [10]: 328 videos from 41 participants
- Convergence curves for DRML, AlexNet, and LCN.

▲ DISFA [11]: 27 videos from 27 participants
- Convergence curves for DRML, AlexNet, and LCN.

Regions v.s. Patches

▲ Importance between regions and patches
- Active region: Magnitude of per-pixel gradients with respect to a particular AU, ie, saliency map [8].
- Active patch: L2 norm for pair-AU classifier, ie, [3] and APL [4].

▲ Learned regions and patches
- The face images were selected manually from the CK+ dataset [8].

Experiments

▲ Settings
- Prediction: 6D feature + Linear SVM.
- Metrics: P1 score and AUC.

▲ Scenarios
- Within-dataset: 5-fold partition for training/valid sets.
- Cross-dataset: 7-fold partition.

▲ Convergence of DRML
- Convergence curves for DRML and AlexNet for Facial Action Unit Detection.

▲ BIP4D [10]: 328 videos from 41 participants
- Convergence curves for DRML, AlexNet, and LCN.

▲ DISFA [11]: 27 videos from 27 participants
- Convergence curves for DRML, AlexNet, and LCN.