REVIEW

MEMBER NAME
CATHERINE HABIB

Paper Title:

Contrastive Attention Network with Dense Field Estimation for Face Completion

Publication Year 2021

This paper proposes a self-supervised Siamese inference network based on contrastive learning for face completion. They assumed that two identical images with different masks form a positive pair while a negative pair consists of two different images. Contrastive learning aims to maximize (minimize) the similarities of positive pairs (negative pairs) in representations. To deal with geometric variations of face images, they integrated a dense correspondence field into their network, this field binds 2D and 3D surface spaces which can preserve the expression and pose of the input. They further proposed a multi-scale decoder with a novel dual attention fusion module (DAF) that can explore feature interdependencies in spatial and channel dimensions and blend features in missing regions and known regions naturally. This multi-scale architecture was beneficial for the decoder to utilize discriminative representations learned from encoders into images. Training their method had two stages. In the first stage, the inference network is trained through contrastive learning until convergence. And in the next stage, the pre-trained encoder and the decoder are jointly trained with the fusion module. For synthesizing richer texture details and correct semantics, loss functions were used such as the element-wise reconstruction loss and many others. In addition, they employed the identity preserving loss function to ensure that the identity information of the generated images remains unchanged. They conducted their experiments for two main tasks face completion and face verification. Their model was trained and tested many times on different publicly available datasets. They used CelebA, CelebA-HQ, FFHQ, and L2SFO for face completion training and testing. As for face verification, they used the training sets of CelebA and Multi-PIE and the test sets of LFW and IJB-C. Their method was implemented by the Pytorch framework and trained on four NVIDIA TITAN Xp GPUs (12GB). Peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and Fréchet Inception Distance (FID) were used as evaluation metrics for most testing. Their method achieved the best quantitative results in three metrics on all testing sets of CelebA (PSNR: 33.26, SSIM: 0.9769, FID: 0.7981), CelebA-HQ, and FFHQ compared with other state-of-the-art methods, which consists of two image inpainting methods, GMCNN and DFNet, and three image-to-image translation methods: Spade, CycleGAN and CUT. They then conducted experiments for face completion on a real-world masked face dataset (RMFD) and got an F1-Score of 0.0493. The face verification tests on LFW and IJB-C datasets achieved high accuracies as well while using the area under the ROC curve (AUC) as the evaluation metrics in the experiments.

Advantages

Using the self-supervised Siamese inference network helped improve the generalization and robustness of encoders. Together with the dense correspondence field and the novel dual attention fusion module (DAF) clearly showed in their detailed and extensive conducted experiments that the proposed approach not only achieves more appealing results compared with state-of-the-art methods but also improves the performance of masked face recognition dramatically.

Disadvantages

Really heavy and quite expensive process. The method consumes a lot of hardware and memory capacity.