1. Introduction:

- Facial Landmark Detection (FLD) aims to detect coordinates of the predefined landmarks on given facial image.
- Can be helpful in various computer vision applications such as 3D face reconstruction, face identification and emotion recognition.
- FLD is challenging due to high variability in poses, lighting and expressions.
- Existing FLD algorithms are either based on coordinate regression or heatmap regression. None of these methods have focused on robust image augmentation techniques to solve the challenges.

2. Objective:

- This work aims to predict accurate landmarks on faces to learn facial semantic structures effectively and outperforms other state-of-the-art (SOTA) methods.

3. Contributions:

- A new patch-based augmentation technique called Fiducial Focus Augmentation (FiFA) is proposed for FLD task to learn facial semantic structures effectively.
- We employ a Siamese-based training scheme utilizing Deep Canonical Correlation Analysis (DCCA) loss between feature representations of two different views of the same image, that enforces consistent predictions of the landmark for the two views.
- To incorporate virtues of both a Transformer and a CNN, we design a robust Transformer + CNN-based backbone.

4. Methodology:

1. Fiducial Focus Augmentation (FiFA)
   - To maximize the correlation between two different augmented views which can be expressed as
     \[ \text{corr}(f_1(l'), f_2(l'')) = \frac{\text{cov}(f_1(l'), f_2(l''))}{\sqrt{\text{var}(f_1(l')), \text{var}(f_2(l''))}} \]
   - The DCCA loss \( L_{\text{DCCA}} \) is then computed as
     \[ L_{\text{DCCA}} = -\text{corr}(f_1(l'), f_2(l'')) \]

2. Siamese training with DCCA loss [1]
   - To introduce desirable properties of CNN while retaining characteristics of transformer.
   - Anti-aliased CNN [2]: To mitigate the loss of structural information caused by pooling layers.
   - FF-Parser layer [3]: To reduce high-frequency noise produced by hour-glass module.

5. Experimental Details:

- Trained/tested on the various benchmark datasets, i.e., WFLW, 300W, COFW and AFLW.
- Along with the DCCA loss, we employ the standard binary cross entropy (BCE) loss and mean absolute error loss for heatmap and coordinate regression, respectively.
- For evaluation, we used the standard evaluation metrics i.e., Normalized Mean Error (NME) variants (i.e., \( NME_{ic} \), \( NME_{bbox} \), \( NME_{diag} \)), Failure Rate (FR), and Area Under the Curve (AUC).

6. Result Analysis:

- Table 1: Effect of method's components on COFW.
- Table 2: Effect of proposed FiFA augmentation technique and Siamese-based DCCA loss on COFW.
- Table 3: Comparison with state-of-the-art methods on Pascal COFW, 300W and AFLW datasets.

References:


