Cross-pose
Identifying faces across a wide range of poses:

TP-GAN [91], PIM [342], DREAM [22], DA-GAN [344], DR-GAN [227], UV-GAN [43], CAPG-GAN [88], PAMs [160], AbdAlmageed et al. [1], MvDN [107]

Cross-pose: In unconstrained conditions, such as surveillance video, the cameras cannot always capture the frontal face image for every appeared subject. Thus, the captured faces have large pose variation from frontal to profile view. As aforementioned in the face preprocessing section, converting profile face images to the frontal pose is a feasible way for cross-pose face recognition, such as TP-GAN [91], FF-GAN [295], PIM [342], and FNM [177]. However, generating the frontal faces will increase the burden of face recognition systems. Cao et al. [22] alleviate this issue by transforming the representation of a profile face to the frontal view in the feature space. Another problem is that the number of profile faces are much fewer than frontal faces in the training data. Thus, some generative approaches [43, 88, 227, 344] proposed to synthesize identity-preserving faces of arbitrary poses to enrich the training data. Moreover, certain methods [1, 107, 160] developed multiple pose-specific deep models to compute the multi-view face representations.

5.1.2 Specialized architectures.

The aforementioned architectures were initially proposed for general visual tasks. Besides, many works develop specialized architectures for face representation learning. At first, many studies [47, 213, 217, 218] attempted to assemble multiple convolution
networks together for learning multiple local features from a set of facial patches. Given the human face appears with regular arrangement of facial parts (eyes, nose, mouth, etc.), such combination of multiple networks with respect to facial part can be more reliable than a single network. Later, based on Bilinear CNN [134], Chowdhury et al. [34] utilized a bilinear architecture for face representation learning. Besides, Xie et al. [276] designed an end-to-end architecture, namely Comparator Network, to measure the similarity of two sets of a variable number of face images. Similar to the multi- network assembling, Comparator Network employs the local attending to facial parts to boost the set-wise representation learning. Han et al. [75] proposed a contrastive CNN to deal with the task of face verification via generating contrastive kernels for convolution so that the features are adaptive to the input face pair. Kang et al. [109] introduced a pair-wise relational network to capture the relations between a pair of local appearance patches. Further, AFRN [108] improve the pair-wise relational network with the attention mechanism. More recently, FANFace [283] integrates the face representation network and facial landmark localization network, so that the heatmap of landmarks will boost the features for recognition.

Some recent works, such as PFE [199] and DUL [25], proposed to take into account the data uncertainty for modeling deep face representation, in order to address the problem of uncertainty caused by low quality face images.