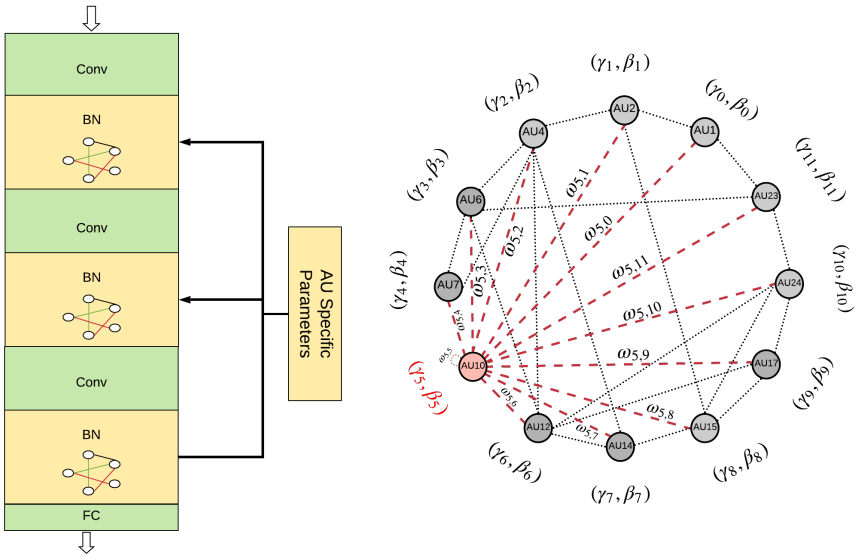

Fig. 1. Framework of proposed method. Each BN layer of the Resnet18 is replaced by its counterpart, calculated as Eq(4) and Eq(5). Each AU in the graph (circles) contains its own parameters (γ, β), and is connected to the others through dotted lines with weight $\omega_{i,j}$. The convolutional layers are shared by all the AUs. For example, if we aim to detect AU10 (red circle) for the input image, the shared convolutional layers and the AU-specific parameters $(\gamma_{10}, \beta_{10})$ are combined to predict the absent or occurrence of AU10, where $(\gamma_{10}, \beta_{10})$ is computed as a weighted combination of the connected AUs in the graph. To make the model's training procedure as fast as the training of ResNet18, a single AU is randomly selected to optimize for each training batch. The graph is constructed using the Pearson Correlation Coefficient (PCC) calculated from the AU labels in the training dataset.