

End2End Occluded Face Recognition by Masking Corrupted Features

汇报人：曹至欣 日期：2021-12-15

1.处理遮挡问题的任务分类

1.1 OFR(occluded face recognition):

目标：将人脸图像与随机部分遮挡进行匹配

与MFR的不同：OFR遮挡位置不确定，目的是学习对于任何位置都可能出现的遮挡具有鲁棒性的特征；MFR遮挡位置已知，可以使用先验知识

1.2 MFR(masked face recognition):

提出背景：疫情

目标：将戴口罩人脸图像与干净人脸图像进行匹配

1.3 PFR(partial face recognition):

目标：将部分人脸图像与普通人脸图像进行匹配

与MFR的不同（如果直接将PFR方法应用到MFR上）：

1. PFR需要一个预先定义的部分人脸作为输入，但这对MFR来说很难检测或者语义定义
2. PFR常常会忽略一些全局信息（例如下巴轮廓）

2.处理遮挡问题的方法分类

2.1 恢复

思路：指首先恢复遮挡的面部部分，然后对恢复的人脸图像进行识别。

例子：一种LSTM自动编码器模型，分为两个部分：第一部分多尺度空间LSTM编码器学习不同尺度的人脸表征，第二部分双通道结构LSTM解码器利用前一部分学到的表示进行人脸重建和遮挡检测。

重难点：保留身份信息的同时需要恢复面部部分

2.2 移除

思路：首先移除被遮挡破坏的特征，然后利用剩余的干净特征进行识别

关键：找到输入人脸图像的遮挡与深层特征之间的对应关系

例子：一种成对差分孪罗网络（PDSN），通过干净人脸和遮挡人脸学习mask字典，处理随机遮挡时组合字典项得到特征丢弃mask与干净人脸相乘，移除掉被遮挡破坏的特征元素。

- 该例子的不足之处：需要额外的遮挡检测模块；训练时间长，训练规模大

3.FROM (face recognition with occlusion masks)

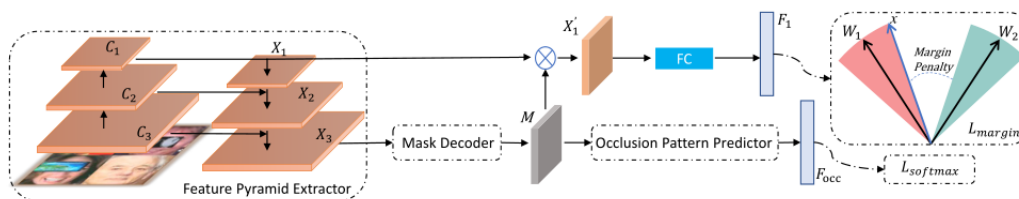


Fig. 2. The pipeline of the proposed FROM. FROM first takes a mini-batch which consists of different random occluded and occlusion-free (not paired) face images as input, and generates a feature pyramid (including X_1, X_2, X_3). Then X_3 is used to decode the masks, which are later applied to X_1 to mask out the corrupted feature elements for the final recognition. We also propose to leverage the occlusion patterns as the extra supervision to guide the feature masks learning. The whole network is trained end-to-end.

3.1 feature pyramid extractor

引入目的：用于获得学习mask的空间感知特征+用于识别的可判别性特征

结构：应用LResnet50E-IR结构，作为模型的主干网络，该网络采用带有横向连接结构的自顶向下结构来构建金字塔特征

输入：未配对的随机遮挡和无遮挡的人脸图像

输出：金字塔特征 X_1 、 X_2 和 X_3 ，其中包含 X_3 包含全局信息和局部信息，被送入mask decoder获得mask用于去除 X_1 被破坏的元素

$$X_2 = \text{conv}(\text{upsample}(\text{conv}(X_1)) + \text{conv}(C_2)), \quad (2)$$

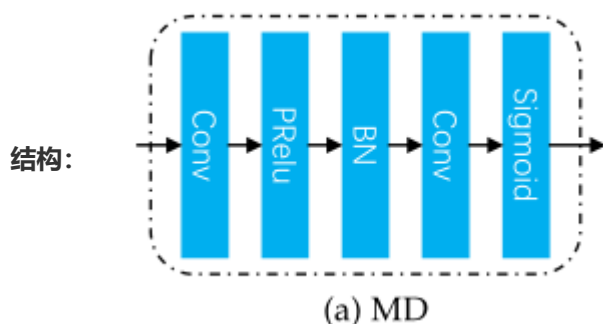
$$X_3 = \text{conv}(\text{upsample}(\text{conv}(X_2)) + \text{conv}(C_3)), \quad (3)$$

3.2 mask decoder

引入目的：输入 X_3 ，获得mask用于去除 X_1 被破坏的元素得到 X_1'

方法：选取深层conv层的特征而不是fc层的特征

- 原因分析：需要得到不依赖于身份的只依赖于遮挡位置的mask，fc层保留了身份信息的同时丢失了空间信息



生成的mask需要达到的目标：

- mask需要与输入图像的遮挡位置相关，通过occlusion pattern predictor实现
- 正确消除掉那些由于遮挡而损坏的特征，通过CosFace Loss监督

实验细节：soft mask VS hard mask？选择soft mask

3.3 occlusion pattern predictor

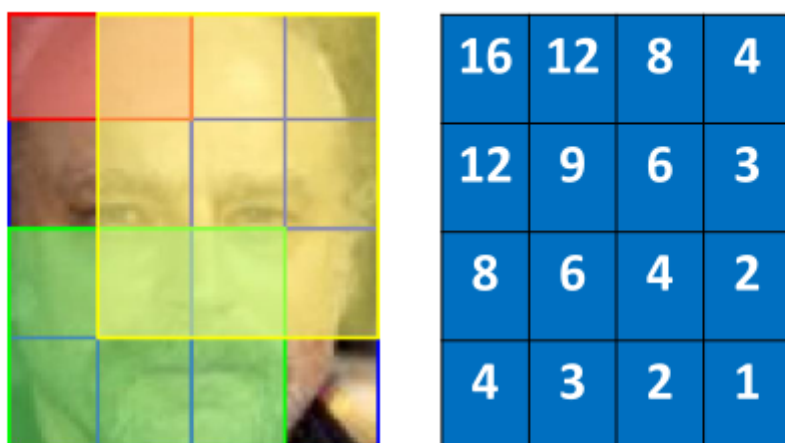


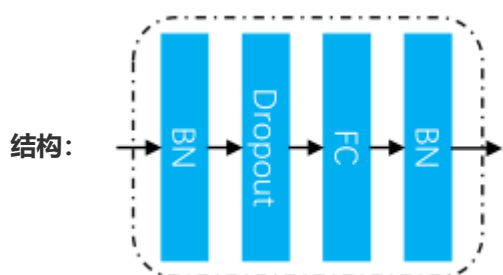
Fig. 4. Examples of proximate occlusion patterns (left) and the number of patterns for each occlusion size (right) when the face image is divided into 4×4 blocks. The value of (i, j) location in the numerical matrix means the number of occlusion patterns with size $i \times j$.

引入目的：监督mask的学习，用于鼓励mask decoder生成与输入人脸的遮挡模式相关的mask

启发思想：一般相邻块在实际应用中具有相同的遮挡状态，eg.如果嘴被遮挡，鼻子很有可能也被遮挡

实现思路：将人脸分成 $K \times K$ 块，用不同大小的矩形覆盖相邻的块，右边的数值矩阵展示了每个不同大小遮挡矩阵能产生的遮挡模式的个数，其中 (i, j) 代表用 $i \times j$ 大小的矩阵覆盖人脸的遮挡模式的数量。最终我们能够获得 $(K \times (K + 1)/2)^2 + 1$ 种遮挡模式，比 $2^{K \times K}$ 少了很多

输入输出：输入mask，输出occlusion feature vector



备注：能够很好的应用于干净人脸识别，即用 0×0 块的矩形去覆盖

具体方法：

1. 描述为分类问题(pattern prediction)

- 在训练阶段，我们得到了每幅图像的遮挡位置。对于每个图像 x_i ，我们通过将其遮挡位置与226个(当 $K=5$ 时)遮挡模式进行匹配来获得其遮挡模式 y_i 。我们的匹配策略是计算遮挡模式和226个参考模式之间的IOU得分，然后选取IOU得分最大的模式作为对应的标签。

$$\mathcal{L}_{softmax} = \frac{1}{N} \sum_{i=1}^N -\log p_i = \frac{1}{N} \sum_{i=1}^N -\log \frac{e^{f_{y_i}}}{\sum_{j=1}^C e^{f_j}},$$

下面展示了输入图像和预测出来的遮挡模式：

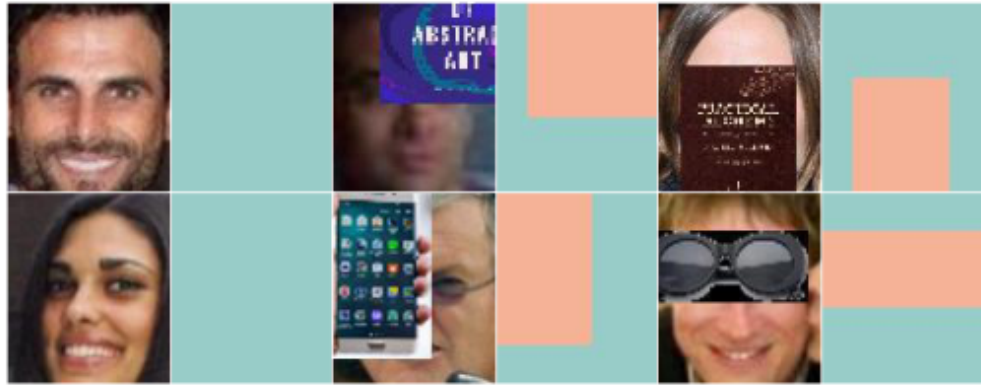


Fig. 5. Examples of input face images and their occlusions predicted by our FROM, where orange color indicates the occluded area. The first row is from Occ-WebFace, and the last row images are from Occ-LFW-1.5. By decoding the correct masks of the corresponding occlusion, corrupted features can be effectively removed.

2. 描述为回归问题(pattern regression)

- 不进行块划分的情况下直接回归遮挡区域的2D坐标，在训练阶段，我们对每一张人脸图像都有遮挡的矩形位置，并直接以左上角和右下角的二维坐标作为label，需要occlusion pattern predictor获得的vector

$$\mathcal{L}_{reg} = \frac{1}{N} \sum_{i=1}^N \|r - f\|_2,$$

3.4 loss

$$\mathcal{L}_{total} = \mathcal{L}_{margin} + \lambda \mathcal{L}_{pred},$$

其中识别人脸使用margin-based softmax loss(CosFace Loss, 其中 $m_1=m_2=0$):

$$\mathcal{L}_{margin} = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s \cdot \delta}}{e^{s \cdot \delta} + \sum_{j \neq y_i}^n e^{s \cos \theta_j}},$$

$$w.r.t. \quad \delta = \cos(m_1 \theta_{y_i} + m_2) - m_3,$$

reference:

Paper:

[1] Ding F, Peng P, Huang Y, et al. Masked Face Recognition with Latent Part Detection[C]// MM '20: The 28th ACM International Conference on Multimedia. ACM, 2020.

[2] Qiu H, Gong D, Li Z, et al. End2End Occluded Face Recognition by Masking Corrupted Features[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, PP(99):1-1.

Blog:

[特征金字塔网络Feature Pyramid Networks - core! - 博客园\(cnblogs.com\)](http://featurepyramidnetworks-core.blogspot.com)