漫画人脸照片识别

李扬波，廖海川，邵永成，钟　杰

摘要：本文针对漫画照片与人脸照片的识别与匹配问题，提出了一种跨模态异质人脸识别方法。该方法包括三个主要步骤：人脸特征表示、解决跨模态问题和设计匹配算法。在处理人脸特征表示时，需要定位人脸的特征点和提取面部特征，但由于漫画和照片的异质表现，传统基于照片的人脸识别方法不适用。在处理跨模态问题时，需要提取一般的特征以防止过拟合单个模态的特征导致引入过多噪声。在设计匹配算法时，利用从照片和漫画提取到的特征进行处理，以判断他们是否是同一个人。本文调研了五种方法，包括 WebCaricature、图片合成法、基于 Facial landmarks 的特征表示、基于重构的方法和基于深度学习的方法。最后，本文以 Rank-1 准确率为评估标准对方法进行了比较分析，提出了我们可能的优化方向和未来研究方向。

1　问题分析

选题 5 为漫画照片与人脸照片的识别与匹配问题。该问题的本质是一个跨模态的异质人脸识别问题，即漫画和真人照片两个模态，其中的人脸是异质表现的。解决这一问题，大致分为三个步骤：人脸特征表示，解决跨模态问题和设计匹配算法。

1.1　人脸特征表示

由于照片和漫画中的人脸是异质表现的，传统人脸识别（基于照片的人脸识别）中的特征点定位和面部特征提取不能很好地直接运用在漫画识别中。在漫画中，除了与照片中一样对主体面部的客观表现，还加入了艺术家的主观印象和绘画风格。这些变量会对人脸特征表示造成很大挑战。

1.2　跨模态问题

对于漫画照片而言，会有面部外观的夸张，因此原本人脸中的特征点会被夸大并出现在不合理的位置而难以被基于照片的人脸识别定位。而漫画有多种绘画风格，导致原本照片和漫画形成的双模态问题就可能变成多模态问题。因此，需要提取一般的特征，防止过拟合单个模态的特征导致引入过多噪声。

1.3　设计匹配算法

利用从照片和漫画提取到的特征进行处理（分类器设计），以判断他们是否是同一个人。

1.4　评测标准

需要计算找出与 Probe 中的图片人物身份相同的 Gallery 图片，返回该图片的名称作为 Probe 图片的匹配结果，赛方计算 Rank-1 准确率。其中，照片与漫画交替作为 Probe 与 Gallery 测试集。

$$Rank-1=\frac{\Sigma_{i=1}^{n}I(G_{i1}=P_i)}{n}$$

2　相关工作调研

在具体的调研的方法实现中，人脸特征表示，解决跨模态问题和匹配算法设计常常结合起来解决。我们一共调研了 5 类方法。

2.1　CNN:VGG-Face

使用以 CNN 为基础的方法，直接利用已有的 VGG-Face 模型, 以相同的流程提取照片和漫画的特征，然后将提取出来的特征输入到传统度量学习方法（如 PCA，KDA）训练的分类器中，以判断照片和漫画是否来自同一个人。由于这两种模态具有很大差异，这种不考虑照片-漫画多模态差异的迁移学习方法，在识别漫画脸部特征时效果不佳，因此对比时具有十分有限的性能。

2.2　图片合成法

将某一模态的图片经过转换生成得到另一模态的图片，然后在同一模态中进行匹配和识别。将人类描述的图片称为 Sketch(如漫画)，则图片合成法在漫画照片人脸识别中的作用就是进行 Sketch-photo 转换或 photo-Sketch 转换，典型的工作如 MRFs（如图 1）和 LLE。这种方法解决了上述跨模态产生差异的问题，但是其缺点在于将多模态的图片直接转化为单模态图片进行特征提取的计算量巨大。尽管可以通过采用 GANs（如图 2）进行优化以降低计算量，然而同样难以训练，且可解释性较差。

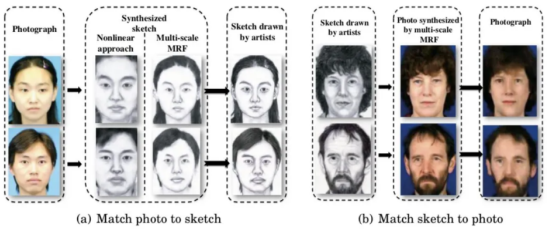

图 1　Sketch-photo 的转换生成

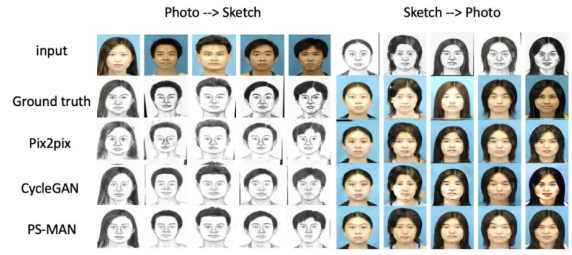

图 2 利用 GANs 进行 photo 和 sketch 的相互转换效果图

2.3 基于 Facial landmarks 的特征提取方法

基于 Facial landmarks 的特征提取方法（如图 3）利用每个 Facial landmark 具有的视角和尺度两维参数，而其具体值由所采用的的特征提取方法决定。基于 Facial landmarks 的特征提取方法以固定的视角和尺度的 landmark 提取照片特征，以不同的视角和尺度的 landmark 提取漫画特征，于是每个面部地标提取到了多个跨模态指标（这是由提取漫画特征时使用了不同视角和尺度导致的）。然后利用跨模态度量学习方法（如使用距离级别的池化）来实现一个照片特征到一组漫画特征的最佳匹配，以减小照片和漫画之间视角和尺度的失调，从而达到跨模态识别的效果。

下面两种特征提取方法与基于 Facial landmarks 的特征提取方法相结合，可以将人脸的 landmark 的特征信息提取，用人脸上代表性部位的特征信息对比来匹配漫画与真实图片。然而由于漫画家不同的创作风格，漫画人脸的某些特征夸张变形严重甚至出现在极不合理的位置，导致面部特征点定位困难，特征难以捕捉，这也是以下两种方法的通病。

2.3.1 特征设计法

特征设计法需要通过人工设计或通过学习得到在不同模态间仍然保持一致的人脸特征，同时这些特征还应满足在不同人脸间的区别度足够高，具体包括 Gabor,SIFT,LBP 等方法。人工设计时，这种方法的缺点在于其需要手动设计特征，虽然基于视觉神经理论，但毕竟是人为设计，难免有想当然，不妥的成分；同时在通过学习得到特征时，该方法严重依赖所给数据库，需要根据提供的数据的特点来进行设计，也就是说设计的特征不适于所有的数据集，泛化性、鲁棒性较差。当数据来源发生改变，如对 RGB 数据设计的特征换成了 Kinect 深度图像，这些特征就不一定适用，因此往往需要重新设计特征。

2.3.2 结合 CNN 特征提取

例如使用 VGG-Face model，这种方法比直接利用 CNN 模型输出特征图的所有特征有更好的性能，比 Webcaricature 只考虑单模态更能处理跨模态的表征差距，因为 CNN 网络强大的特征提取能力，使得这种方法性能较好。只是相对于后文提到的多任务学习方法，有些特征难以学习。

2.4 多任务学习

与 WebCaricature 等单任务学习方法相比，多任务学习方法能同时利用不同的数据训练不同的任务，如以漫画和图片作为训练数据时，多任务模型可以同时执行照片-漫画面部验证，漫画识别和照片识别，从而忽略依赖数据的噪声以学习更一般的特征。又由于该方法整合了不同的任务，所以可以学习到对于某些任务难以学习的特征。

2.4.1 多任务学习方法

多任务学习方法分为两种：基于硬参数共享和基于软参数共享的多任务学习方法。其中，硬参数共享指的是多个任务之间共享网络的同几层隐藏层，只不过在网络的靠近输出部分开始分叉去做不同的任务；而软参数共享则是不同的任务使用不同的网络，但是不同任务的网络参数，采用距离(L1,L2) 等作为约束，鼓励参数相似化。由于不同形态的漫画和照片存在一些共同的面部特征，故使用硬参数共享的多任务学习方法是更好的选择。

2.4.2 搜索方法

多任务学习需要为每个任务设置权重，而搜索最优权重的方法主要为静态搜索与动态搜索。静态搜索中，利用实验方法手动搜索效率低，利用贪心搜索方法则费时。利用动态搜索时，若利用网络的总损失更新任务的动态权重时，容易陷入简单任务的过度训练和困难任务的不足训练。因此采用计算各个任务的损失并为具有较大损失的任务设置较大的任务权重，以重点学习困难任务。具体地，在漫画人脸照片匹配问题中，可以将漫画照片与人脸照片匹配看作主任务，将对漫画照片人物 ID 与真实人脸照片人物 ID 进行识别匹配看作另外两个子任务。将这三个任务通过最后一个共享的隐含层参数进行联系，其利用提取的任务之间的公共信息来学习任务权重。该共享隐藏层与动态权重学习模块（一个具有 softmax 归一化的全连接层）相连，使其生成三个任务的动态权值。该动态权重损失模块还能将各任务的损失和对应的动态权值，输入到一个新的损失函数中，以学习驱动网络专注于训练困难任务的动态权值，这样就改进了漫画人脸匹配的学习参数，使之能够有较好的效果。下图（图 4）是上述自动学习生成动态权重的跨模态照片漫画识别动态多任务学习网络框架。

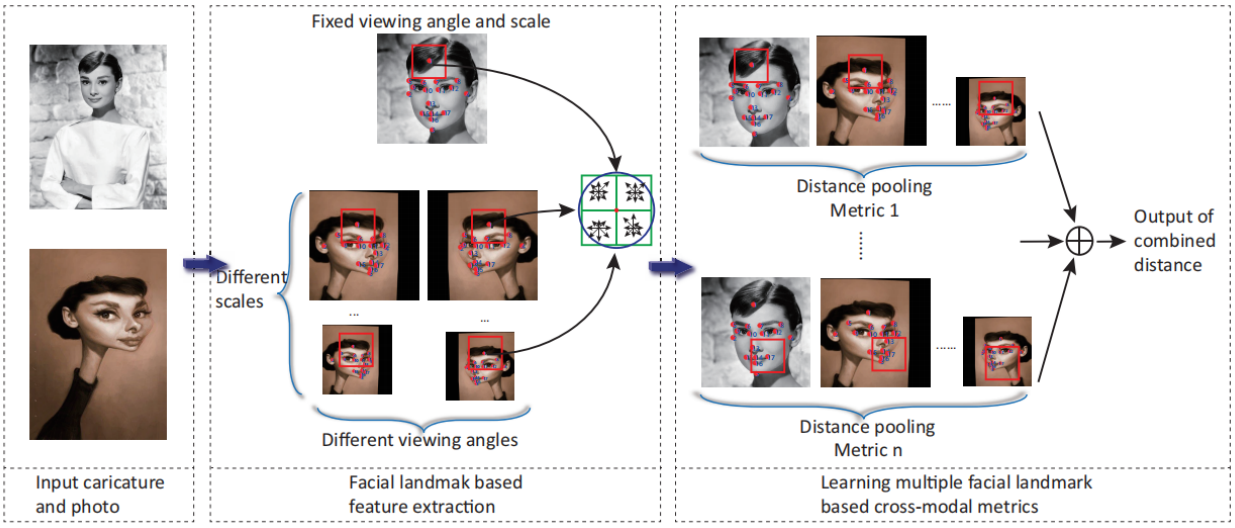

图 3　基于 Facial landmarks 的特征提取方法流程图

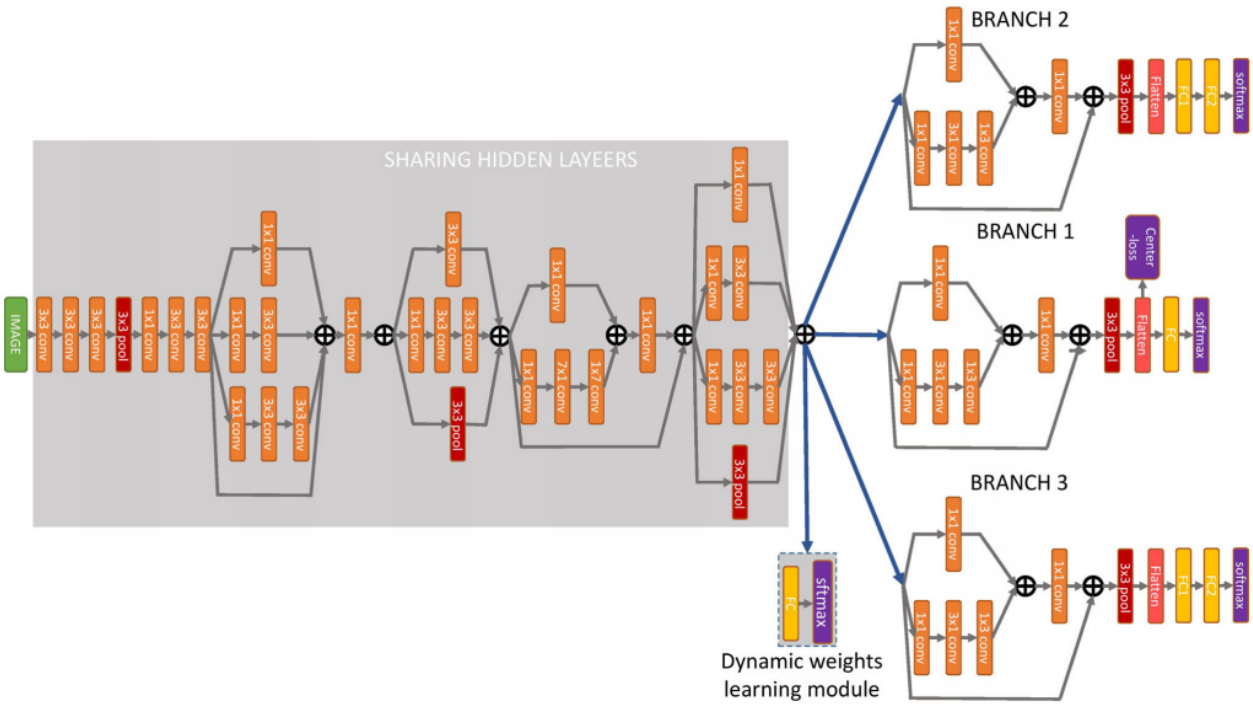

图 4　自动学习生成动态权重的跨模态照片漫画识别动态多任务学习网络框架图

3 不同方法的评估结果

3.1 数据集 Webcaricature

Method	C2P		P2C	
	Rank-1 (%)	Rank-10 (%)	Rank-1 (%)	Rank-10 (%)
Euc	13.38±1.10	38.40±1.92	9.04±0.80	29.63±1.35
PCA	15.63±0.82	43.48±1.69	12.47±1.14	40.13±1.53
KDA	19.32±1.36	56.77±1.58	18.92±1.35	57.19±2.61
KissME	15.16±1.63	43.95±2.13	13.30±1.18	43.63±1.64
ITML	15.25±3.07	46.39±6.46	16.48±1.77	49.88±2.29
LMNN	17.92±0.86	50.58±1.72	15.90±1.73	48.08±1.95
CCA	10.84±0.78	40.76±1.08	10.73±0.94	41.12±1.87
MvDA	4.77±0.74	27.73±1.90	4.71±0.87	27.19±2.55
CSR	**25.18±1.39**	60.95±1.20	23.36±1.47	60.27±1.97
KCSR	24.87±1.50	**61.57±1.37**	**23.42±1.57**	**60.95±2.34**

图 5 不同学习方法在 C2P 与 P2C 中的学习效果

C2P		
Method	Rank-1 (%)	Rank-10 (%)
SIFT-Land-KCSR	24.87±1.50	61.57±1.37
VGG-Eye-PCA	35.07±1.84	71.64±1.32
VGG-Eye-KCSR	39.76±1.60	75.38±1.34
VGG-Box-PCA	49.89±1.97	84.21±1.08
VGG-Box-KCSR	**55.41±1.41**	**87.00±0.92**
P2C		
Method	Rank-1 (%)	Rank-10 (%)
SIFT-Land-KCSR	23.42±1.57	60.95±2.34
VGG-Eye-PCA	36.18±3.24	68.95±3.25
VGG-Eye-KCSR	40.67±3.61	75.77±2.63
VGG-Box-PCA	50.59±2.37	82.15±1.31
VGG-Box-KCSR	**55.53±2.17**	**86.86±1.42**

图 6 不同特征提取方法在 C2P 与 P2C 中的效果

Method	Rank-1 (%)	Rank-10 (%)
SIFT-Land-KCSR	24.87 ± 1.50	61.57 ± 1.37
VGG-Eye-PCA	35.07 ± 1.84	71.64 ± 1.32
VGG-Eye-KCSR	39.76 ± 1.60	75.38 ± 1.34
VGG-Box-PCA	49.89 ± 1.97	84.21 ± 1.08
VGG-Box-KCSR	55.41 ± 1.41	87.00 ± 0.92
Navie Dynamic	86.00 ± 1.70	98.21 ± 1.08
Ours (Single-verif)	85.55 ± 1.30	96.31 ± 0.08
Ours (Dynamic MTL)	**87.30 ± 1.20**	**99.21 ± 1.07**

图 7 不同特征提取方法在 C2P 中的效果 (与多任务学习对比)

Method	Rank-1 (%)	Rank-10 (%)
SIFT-Land-KCSR	23.42 ± 1.57	69.95 ± 2.34
VGG-Eye-PCA	36.18 ± 3.24	68.95 ± 3.25
VGG-Eye-KCSR	40.67 ± 3.61	75.77 ± 2.63
VGG-Box-PCA	50.59 ± 2.37	82.15 ± 1.31
VGG-Box-KCSR	55.53 ± 2.17	86.86 ± 1.42
Navie Dynamic	82.80 ± 1.60	97.81 ± 0.88
Ours (Single-verif)	81.70 ± 2.60	95.25 ± 1.08
Ours (Dynamic MTL)	**84.00 ± 1.60**	**99.01 ± 1.2**

图 8 不同特征提取方法在 P2C 中的效果 (与多任务学习对比)

3.2 数据集 CaVI

Method	Verification	Photo identification	Caricature identification	V2C	C2V
CaVINet	91.06	94.50	85.09	–	–
CaVINet(TW)	84.32	85.16	86.02	–	–
CaVINet(w/o ortho)	86.01	93.46	80.43	–	–
CaVINet(shared)	88.59	90.56	81.23	–	–
CaVINet(visual)	88.58	92.16	83.36	–	–
Navie Dynamic	93.80	97.60	75.80	61.90	62.80
Ours (Single-verif)	92.46	–	–	–	–
Ours (Single-visual)	–	98.10	-	-	41.80
Ours (Single-cari)	–	–	78.20	53.60	–
Ours (Dynamic MTL)	**94.92**	**98.35**	**85.61**	**80.04**	**64.39**

图 9 不同特征提取方法在 CaVI 数据集中的效果

4 预选方法

通过以上调查，我们学习小组拟决定采取基于卷积神经网络（CNN）的特征提取方式，采用多任务学习方法完成此项任务。

参考文献

[1] HUO J, LI W, SHI Y, et al. Webcaricature: a benchmark for caricature recognition[Z]. 2018.

[2] HUO J, GAO Y, SHI Y, et al. Variation robust cross-modal metric learning for caricature recognition[C]//New York, NY, USA: Association for Computing Machinery, 2017.

[3] MING Z, BURIE J C, LUQMAN M M. Cross-modal photo-caricature face recognition based on dynamic multi-task learning: volume 24[M]. Berlin, Heidelberg: Springer-Verlag, 2021.

[4] WANG X, TANG X. Face photo-sketch synthesis and recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009(11): 31.