

# 可解释性主观情绪特征的情感传达建模

王心畅 律睿敏\* 卞则康\*

江南大学

## Abstract

中国诗词常会寓情于景，其实文学艺术作品具有通性，无论是书法作品还是水墨画、简笔画，作者都会通过艺术作品传达某种或者某些情感。文学艺术作品的意象通常是抽象的，其主观性难以使用客观性的指标去评价。本文通过构建一个主观情绪特征的情感传达模型，将简笔画进行7种情绪的分类。受到计算机视觉中感知损失的启发，本文通过PCA降维与使用VGG16预训练模型进行降维处理，探索简笔画哪些特征与最终情感表达具有相关性，然后通过自定义主观情感特征，使用FPN金字塔情感分类建立一个可解释性感知情感分类模型，可以根据输入的简笔画进行7种情感分类。同时，我们对比不使用定义的感知情感特征，而是直接将简笔画以图片的形式，映射到视觉大模型的特征空间，转化为计算机视觉领域的传统分类问题，测试情感分类的准确性；另一方面，本次与使用预训练模型、使用AI工具（包括GPT、Kimi）进行分类对比，验证了本次构建的可解释性感知情感分类模型的有效性，本次模型构建的整体流程见图1。本次已经开源建模代码与相应的数据集 <https://github.com/ccandtt/MutualMediaTechnology>

## 1. 引言

在深度学习出现之前，大多数方法采用传统的机器学习方法进行字符识别。例如，使用 K-Means 进行字

符分类，使用带有字符骨骼、边缘和笔画信息的 SVM 进行分类，采用多尺度特征提取方法，Gao 等人 [14] 提出了一种基于检索的快速识别解决方案。然而，由于需要大量的人工操作和机器学习方法的性能有限，传统的机器学习方法在书法字符识别中往往表现出较弱的泛化和严重的过拟合。

随着深度学习的发展及其强大的表现能力，它在书法字符识别中得到了应用。例如，Gao 等人 [7] 使用深度卷积神经网络对书法图像进行特征提取，并通过实验验证了其有效性。Xu 等人 [15] 分别使用 GoogleNet Inceptionv3 和 ResNet50 来识别书法作品的风格和字符形式。Mai et al. [14] 改进了 DenseNet [4]，并将其与数据增强方法相结合，增强了模型的鲁棒性和泛化能力。

中国书法字体识别与书法字符识别一致，在深度学习出现之前，字体样式识别主要依赖于机器学习方法。然而，这些方法通常需要手动选择特征，从而限制了模型的通用性。早期的方法包括手动特征选择过程，例如 Gabor 滤波器 [16]、小波变换 [2] 和局部二进制模式（LBP）[6]；然而，这些方法在稳健性方面存在缺陷。

随着深度学习的兴起，神经网络在特征提取方面表现出了出色的性能，有时甚至超过了传统的机器学习算法。例如，Wang 等人 [13] 使用基于补丁的 CNN 模型进行中文字体识别的准确率为 53%。然而，这些方法主要集中在孤立汉字的字体样式识别上。

对于由许多连接笔画组成的字符符号（如阿拉伯语），字符分割的挑战会增加。为了解决这个问

\*Equal contribution

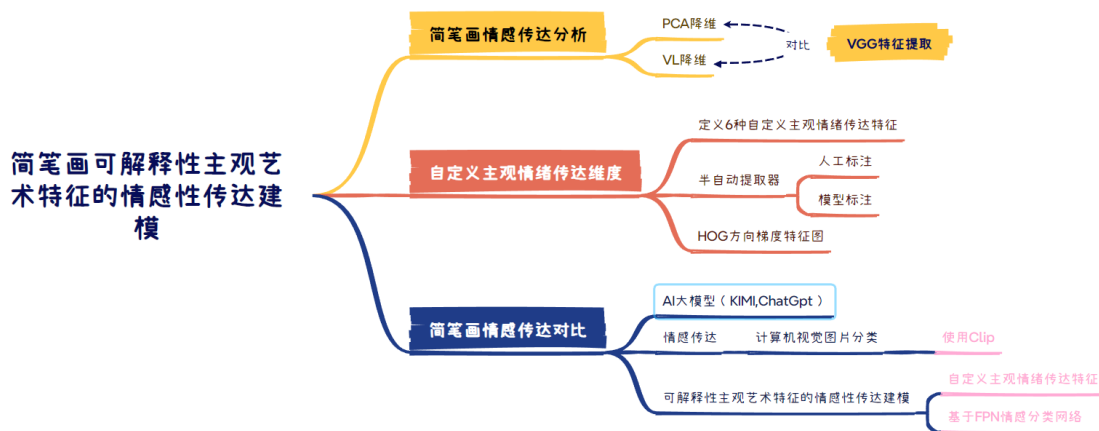


Figure 1. 可解释性主观艺术特征的情感性传达workflow

题，Slimane 等人 [8]提出了一种基于滑动窗口的方法，该方法通过在图像上移动滑动窗口来消除字符分割的需要，并采用高斯混合模型（GMM）进行字体样式分类。然而，这些方法通常需要复杂的图像预处理才能获得有效的分类特征，这并不能保证提取的特征对其他或未来的数据集有效。

在深度学习领域，神经网络在特征提取方面表现出色。Tao 等人 [12]将二维长短期记忆（2DLSTM）模型与主成分分析相结合，实现了高达 97.77% 的识别率，并表现出更大的灵活性和稳健性。在其他研究中，Lee 和 Ding [5]使用自动编码器进行特征提取，与 K-NN 等传统机器学习方法相比，实现了卓越的识别性能。此外，Tang [11]观察到汉字的骨架信息可能是一个关键的分类指标，并提出了使用骨架内核。使用 VGG19 [9] 网络的实验结果表明，使用骨架内核可以将识别准确率提高约 10%。风格和-content 监督（SCS）网络 [10]采用两个独立全连接的层分支来提取汉字的字体样式和内容特征，通过双线性模型的组合实现了 88.06% 的识别准确率。

深度学习方法消除了对复杂预处理步骤和手动特征工程流程的需求。模型可以根据训练数据集自动

学习特征，确保使用最合适的特征。由 Li 等人提出的 SwordNet 是第一个使用跳过连接和全局平均池化（GAP）来增强鲁棒性的深度学习方法。实验结果表明，与现有的深度学习方法相比，它对不同大小的数据集表现出更好的识别准确率，并且也能识别英文文本。

情绪心理学给出了其中七种基本情绪的定义，包括恐惧、中立、惊讶、悲伤、愤怒、高兴、厌恶，并且认为其具有跨文化、跨种族的普遍性，是人类在面对不同刺激和情境时产生的普遍心理反应，如图 ??。1993年Paul Ekman就讨论了面部表情与情绪相关关系 [3]，探讨了面部表情通常传达的信息、情绪是否可以没有面部表情、情绪的面部表情是否可以没有情绪本身，以及个体在情绪的面部表达上的差异。而加上深度学习技术在计算机视觉方面的应用，Augng等人 [1]通过卷积神经网络（CNN）进行基于图像的人脸情感识别并且取得了较好的效果。

人脸能够进行识别是基于人的面部表情可以传达情绪，受此启发，本文不同于人脸情绪分类任务，使用简笔画作为主要分析对象，进行可解释性主观艺术特征的情感性传达建模，通过自定义具有可解释性的感

知特征以及针对本次任务构建的多层语义情感分类模型（以下简称为，**MS-SC**模型）。

## 2. 本次工作描述与定义

**任务描述** 本文通过对简笔画进行可解释性主观艺术特征的情感性传达建模，能够对简笔画的情绪传达做出正确的分类判断。定义模型检测的简笔画为 $X$ ，情绪定义为7种情绪，分别是fear(恐惧)、neutral(中立)、surprise(惊讶)、sadness(悲伤)、anger(愤怒)、happiness(快乐)、disgust(厌恶)，最终目标是通过模型检测简笔画的传达情绪  $E_i$ 。

$$f: X \rightarrow \mathcal{E},$$

其中  $f(X) = E_i$ ,  $E_i \in \mathcal{E}$ ，表示模型根据输入的简笔画  $X$  所预测的情绪类别。具体目标是最小化情绪预测误差，形式化为：

$$\min \mathbb{E}[L(E_{\text{true}}, f(X))],$$

其中：

- $E_{\text{true}}$  表示简笔画的真实情绪标签；
- $f(X)$  表示模型对输入  $X$  的预测情绪；
- $L$  是定义在真实情绪与预测情绪之间的损失函数，例如交叉熵损失。

## 3. 可解释性感知特征的定义

区别于人脸等具有丰富细节特征的数据，简笔画或书法等艺术作品的细节丰富度较低，直接将原始数据集作为训练集时，深度学习模型可能难以从中有效学习到简笔画与情绪之间的抽象关系。为了弥补这一不足，本文使用 **HOG**（方向梯度直方图）特征提取来增强图像的细节表达。通过这种方式，**HOG** 特征可以有效地捕捉到简笔画或书法作品中的关键形状和纹理信息，提升模型对这些抽象特征的识别能力。

在情绪分类任务中，**HOG** 特征作为额外的输入，可以帮助模型更好地理解图像中的边缘、结构及纹理信息，特别是在图像细节较为简单的情况下。通过增强图像的梯度信息，**HOG** 提取的特征不仅提升了对图像中局部变化的敏感度，还提高了模型对光照、旋转等变化的鲁棒性，从而有助于提升情绪分类任务的性能。

此外，结合 **HOG** 特征与深度学习模型（如 CNN）一起使用，还可以利用深度学习模型自动学习图像的高级特征，从而在不完全依赖手动特征提取的情况下，进一步优化分类结果。通过这种方式，**HOG** 特征成为了一种有效的补充，为深度学习模型提供了额外的支持，特别是在处理简笔画、书法等细节相对较少的艺术作品时。

### 3.1. HOG梯度特征增强

方向梯度直方图（**HOG**）是一种基于梯度方向分布的特征提取方法。通过统计图像局部区域中像素梯度方向的分布，**HOG** 能够有效描述边缘和纹理信息，其特征具有对光照和对比度变化的鲁棒性。**HOG** 的核心步骤包括梯度计算、方向直方图构建、归一化以及特征向量生成。

对每个像素，根据公式 1 计算一阶差分计算其水平方向 ( $G_x$ ) 和垂直方向 ( $G_y$ ) 的梯度值：

$$\begin{aligned} G_x &= I(x+1, y) - I(x-1, y) \\ G_y &= I(x, y+1) - I(x, y-1) \end{aligned} \quad (1)$$

其中， $I(x, y)$  表示像素点  $(x, y)$  的强度值。

根据  $G_x$  和  $G_y$ ，可计算梯度幅值  $M$  和梯度方向  $\theta$ ：

$$M = \sqrt{G_x^2 + G_y^2}, \quad \theta = \arctan\left(\frac{G_y}{G_x}\right) \quad (2)$$

将图像划分为多个单元格（Cell），每个单元格内统计梯度方向的直方图。方向被量化为  $n$  个区间（Bins），通常  $n = 9$ 。直方图中每个 Bin 的值是该方

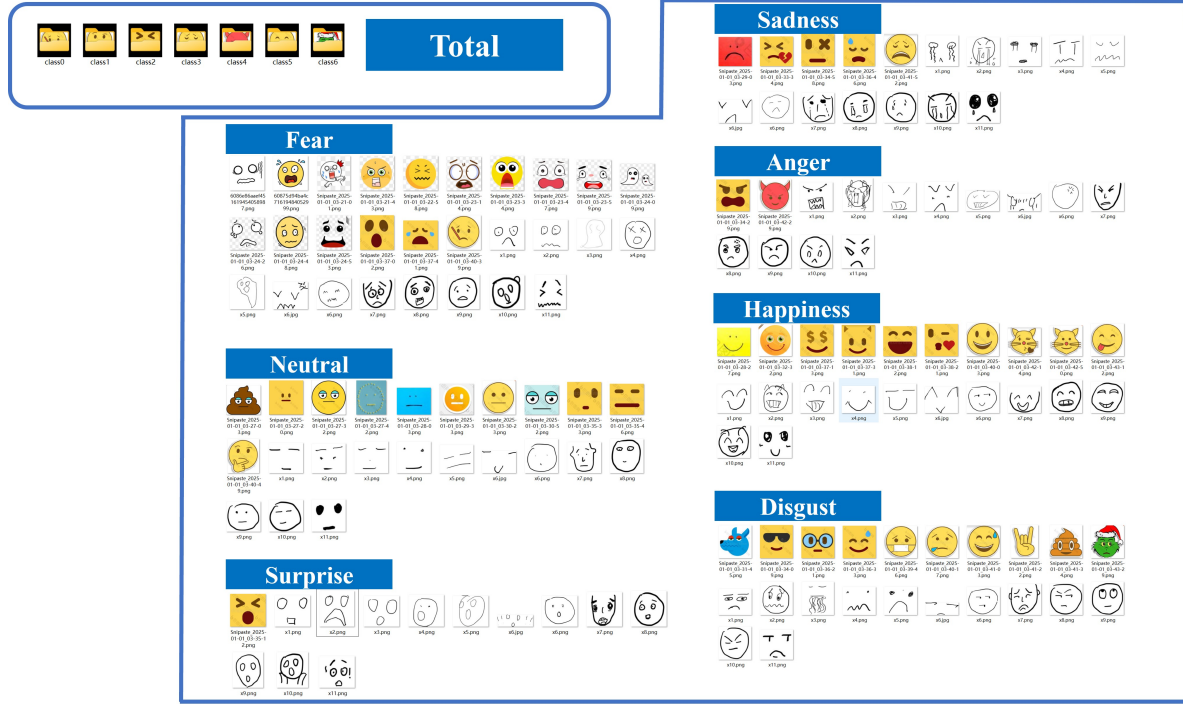


Figure 2. 扩充后简笔画数据集（节选）

向上的梯度幅值之和:

$$h[b] = \sum_{(x,y) \in \text{cell}} M(x,y) \cdot \delta(b, \theta(x,y)) \quad (3)$$

其中,  $\delta(b, \theta)$  是一个指示函数, 用于判断梯度方向  $\theta(x, y)$  是否属于第  $b$  个方向区间。

为了提高特征对光照和对比度变化的鲁棒性, 将相邻的单元格组合成一个块 (Block), 并对块内的梯度直方图进行归一化处理。归一化公式为:

$$h_{\text{norm}} = \frac{h}{\sqrt{\|h\|_2^2 + \epsilon^2}} \quad (4)$$

其中,  $\epsilon$  是一个小常数 (如  $\epsilon = 10^{-5}$ ), 用于防止分母为 0。

将所有块的归一化梯度直方图按顺序拼接, 形成整个图像的特征向量:

$$\mathbf{H} = [h_{\text{norm}}^1, h_{\text{norm}}^2, \dots, h_{\text{norm}}^k] \quad (5)$$

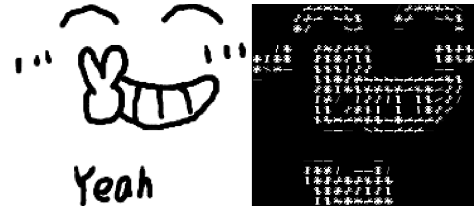


Figure 3. 经过HOG增强的特征图

其中,  $k$  为块的总数。

## 3.2. 自定义主观情绪传达特征

### 3.2.1 预训练模型进行数据降维

在自定义主观情绪化特征之前, 我直接使用预训练模型进行特征提取。传统的数据降维方法大多是通过PCA降维、MDS降维、tSNE降维的方式, 其实是从纯数据分析的角度, 根据像素值大小进行降维处理, 缺少了高级语义。以PCA降维为例, 其实是将原始图像进行重塑成一维数据, 然后按照列向量进行降维, 但是重塑成一维向量这个过程打破了原有图像的空间信息与语义信息, 使得仅利用像素值这一数值信息,



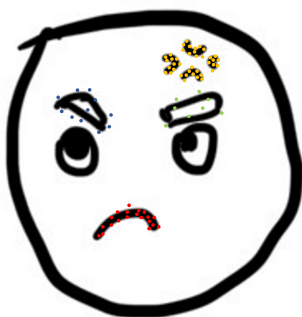


Figure 4. 特征相关性

难以真正探索简笔画图像中传达情绪的语义特征，所以本文并没有采取传统的降维方式，而是使用计算机视觉中常见的视觉预训练模型。

**可视化特征分析** 首先，我简单构造一个使用VGG作为backbone的分类网络，直接将简笔画图像映射到VGG特征空间中，然后添加几层网络作为训练层不进行冻结操作，将情感传达任务转成计算机视觉中的多分类任务。在测试阶段，将倒数第二层的输出特征之间进行可视化，以探求哪些特征与最终的情感分类有直接相关关系。

如图，我们观察到某些特殊符号以及部位比如嘴巴、眼睛，与简笔画的情绪传达具有直接关系。经过统计整理，大致分为“A-嘴巴”、“B-眼睛”、“C-特殊符号”，这样子的结果与客观事实与正常认知相复合，众所周知面部表情主要也是通过眼睛、嘴巴等五官进行传达，通过VGG降维分析，我接下来从六个维度自定义了主观情绪传达特征。

### 3.2.2 主观情绪传达特征的自定义

由于简笔画中的某些符号可以表示某一种情绪，本次我定义了六种在简笔画表达情绪中出现频率最高的特征分别是上弧嘴角、下弧嘴角、感叹号、爆筋、弯眉、笑眼，并且使用One-hot编码，为每个简笔画作品 $X_i$ 新增6个特征，记为 $e = 6$ ，每个维度的数值大小为 $[-1, 1]$ ，其中1表示待检测的简笔画中在该维度上非常复合描述，-1表示与该维度描述特征相反，数



Figure 5. 自定义特征

值在0附近则表示该维度信息传达不明显，具体可见图6。

Table 1. 自定义主观情绪传达维度描述

维度	描述
上弧型嘴角	嘴角是为“上弧”形状，通常可以传达高兴的情绪，反面是“下弧”嘴角。如果有些简笔画中的嘴角是类似“直线”，那么此维度的数值可以取0左右。
下弧型嘴角	嘴角是为“下弧”形状，通常可以传达失落、难过、生气的情绪，反面是“上弧”形状的嘴角。如果有些简笔画中的嘴角是类似“直线”，那么此维度的数值可以取0左右。
感叹号	类似标点符号中感叹号的形状，通常可以传达震惊、惊讶等情绪。
爆筋	表情包中有相关的符号，由3-4个红色的弯曲线条组成，通常传达生气、愤怒的情绪。
弯眉	通常出现在笑脸上
笑眼	简笔画眼部笑眯眯、或者有月牙形笑眼等，总之眼部情绪为开心、笑意。

首先，原始数据集 $X$ 的样本数为 $m$ ，特征数为 $n$ ，手动标注的比例为

$$p(p \in [0, \frac{1}{2}])$$

，对 $m * p$ 的样本进行人工数据标注。同时，自行搜集互联网上表情包或者简笔画，进行扩充数据集，整理成数据集 $Y$ ，其中 $Y$ 的特征数量为 $X$ 的特征数 $n + e$ 。

## 4. 可解释性主观艺术特征的情感性传达建模

由于简笔画等艺术形式的细节丰富度不高，本次采用双支网络进行特征提取，上半支网络构建自定义主








	上扬嘴角	下弧嘴角	感叹号	爆筋	弯眉	笑眼
	0.23	0.24	0.08	-0.16	-0.19	-0.07
	0.18	0.01	0.19	-0.26	-0.36	-0.18
	-0.18	0.18	0.24	0.13	-0.2	-0.06
	-0.14	0.25	0.13	0.09	0.01	-0.1
	0.07	-0.2	0.19	0.16	-0.15	-0.19
	0.2	-0.24	-0.07	0.07	0.01	0.42
	-0.12	0.28	0.07	0.12	-0.22	-0.17

Figure 6. 7维可视化展示

观情绪传达特征提取器，通过半自动提取自定义的主观情绪表达特征，与下半支采用基于FPN金字塔结构的网络输出结果进行融合，使用融合特征来对图像进行情感分类任务，接下来详细介绍两个网络，网络结构见图 7。

#### 4.1. 半自动特征提取器

上述已经介绍，我们首先对 $m * p$ 份样本进行手动人工数据标注，使用标注后的数据构造一个回归任务，预测6个维度的数值，进行自动化标注。具体而言，首先手动训练一个特征检测模型，设计一个检测器，使用分割的局部特征作为训练集，作为自定义主观情绪传达特征提取器(Custom Subjective Emotion Communication Feature Extractor),以下简称为SE-Extractor。

#### 4.2. 基于FPN金字塔结构的情感分类网络

如 7，上半支网络主要提取自定义主观情绪传达特征，接下来重点介绍网络的下半支部分，即基于FPN金字塔结构的简笔画情感分类网络。

使用HOG数据增强 前文已经介绍如何如何如何进行HOG梯度增强，生成的HOG特征记为  $\mathbf{H}$ ，其维度为  $d_{\text{HOG}}$ ，将HOG特征图与原始输入图像在channel进

行堆叠生成增强特征图 $X_{\text{aug}}$ ，并将增强特征图作为接下来FPN网络的输入,见公式 6。

$$\mathbf{X}_{\text{aug}} = \text{Concat}(\mathbf{X}, \mathbf{H}) \quad (6)$$

其中  $\mathbf{X}$  为原始图像， $\mathbf{H}$  为 HOG 特征。此时的输入尺寸为  $128 \times 128 \times c$ ，其中  $c = 3 + \text{HOG 通道数}$ 。

**FPN网络的多尺度特征提取** FPN (Feature Pyramid Network) 通过自顶向下的特征融合策略，有效结合深层高语义信息与浅层高分辨率细节，生成多尺度的特征图。FPN的输出特征可以表示为：

$$\{P_2, P_3, P_4\}$$

其中， $P_2$  到  $P_4$  分别为FPN网络不同层级的输出特征图，随着层次加深，特征图的分辨率逐渐降低，而语义信息逐渐增强。

**多尺度特征的提取与融合** 每个特征图  $P_i$  的计算公式为：

$$P_i = \text{Conv}(U(P_{i+1}) + L_i) \quad (7)$$

其中， $U(\cdot)$  表示上采样操作， $L_i$  表示从CNN主干网络中提取的特征， $\text{Conv}(\cdot)$  为卷积操作。

#### 4.3. 双支网络集成

最终将两个分支网络进行融合，使用融合的特征进行分类。

首先对每个特征图 $P_i$ 进行Flatten操作,同时融合 $F_S E$ 特征，并输入到接下来的分类网络中进行最终的情感分类。

$$\mathbf{F}_{P_i} = \text{Flatten}(P_i), \quad i = 2, 3, 4 \quad (8)$$

$$\mathbf{F} = \text{Concat}(\mathbf{F}_{P_2}, \mathbf{F}_{P_3}, \mathbf{F}_{P_4}, \mathbf{F}_{SE}) \quad (9)$$

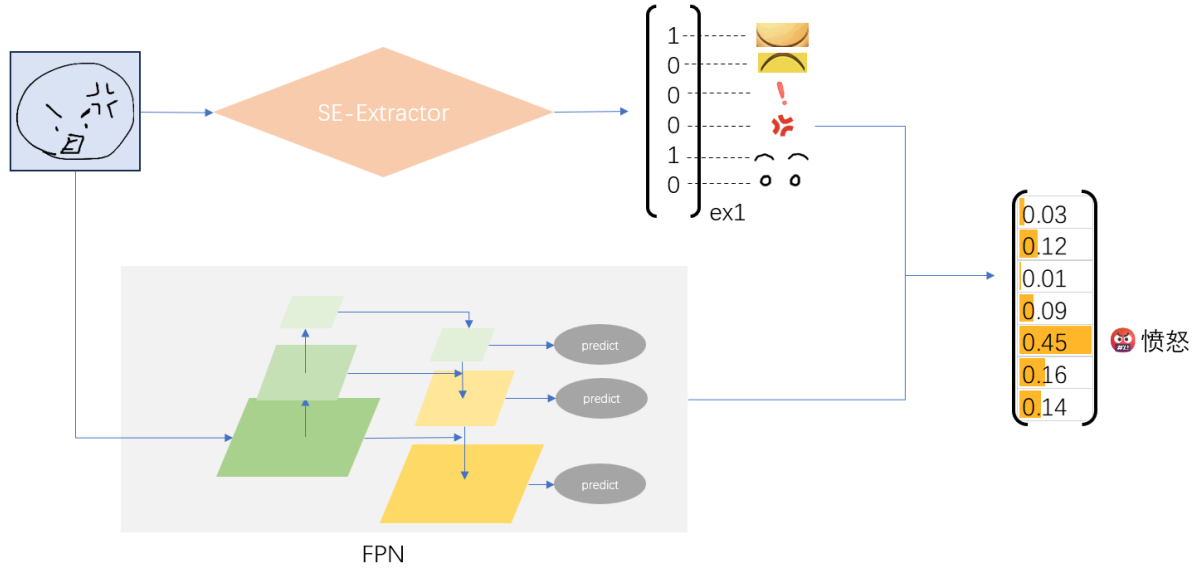


Figure 7. 可解释性主观艺术特征的情感性传达建模.

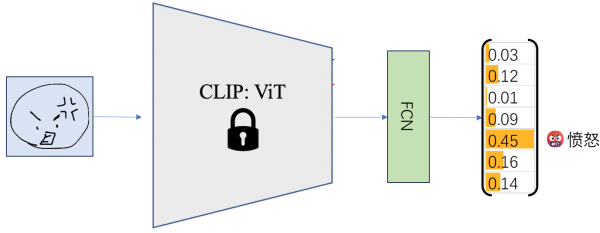


Figure 8. 基于预训练模型的情感分类



Figure 9. AI大模型判断简笔画的情感传达

$$\mathbf{F}' = \text{FC}(\mathbf{F}) \quad (10)$$

$$\mathbf{y} = \text{Softmax}(\mathbf{F}') \quad (11)$$

## 5. 实验

### 5.1. 数据集

### 5.2. 基于预训练模型的情感分类

为验证本次模型的有效性，首先我们将图像作为输入，使用预训练模型进行特征空间映射，然后训练一个线性层进行情感分类。如图 8, 本次在进行对比实验

时，我们基于CLIP进行情感分类，分类的准确率仅可达到23.81%，

### 5.3. 基于AI大模型的情感传达问答

因为目前AI大模型可以进行读图，故此使用AI大模型进行该任务的测试，但是发现AI大模型无法处理这种抽象的情感传达，其回答明确表示需要能够表现人类情绪的面部表情，而不是简笔画。

### 5.4. 实验结果

实验结果见表 2, 可以如果仅仅将本次任务转为计算机视觉的图像分类，准确率会很低，由于简笔画的细节丰富程度不够，但是通过本文HOG数据增强、

	Method1	Method2	Method3
train acc.	/	21.65%	72.61
test acc.	19.1	23.81%	55.05

Table 2. 不同方法的对比。其中Method1是指基于AI大模型的情感传达问答，Method2是指基于预训练模型的情感分类，Method3是使用本文提出的方法进行该任务。其中Method1仅有测试结果。<sup>1</sup>

主观性特征定义、基于FPN金字塔情感分类模型的构造后，准确率有所提升，但是可能由于本次训练集不多，每个类别仅有50张简笔画图片，准确率仍然较低。

## 6. 总结与展望

本文提出了一个新的任务，使用深度学习技术对艺术性图片（简笔画）进行情感性传达建模，并与直接使用预训练模型进行分类、使用AI大模型进行测试对比，发现本文提出的模型确实具有良好的表现效果，但是可能由于数据集规模较少、定义的主观情感维度较少，准确率其实并没有很高，后续可以扩大规模继续研究。

## References

- [1] Erlangga Satrio Agung, Achmad Pratama Rifai, and Titis Wijayanto. Image-based facial emotion recognition using convolutional neural network on emognition dataset. *Scientific Reports*, 14(14429):1–14, 2024.
- [2] Xiaoqing Ding, Li Chen, and Tao Wu. Character independent font recognition on a single chinese character. *IEEE Transactions on pattern analysis and machine intelligence*, 29(2):195–204, 2007.
- [3] Paul Ekman. Facial expression and emotion. *American Psychologist*, 48(4):384–392, 1993.
- [4] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. arxiv 2016. *arXiv preprint arXiv:1608.06993*, 1608, 2018.
- [5] Chi-Chang Lee and Jian-Jiun Ding. Automatic chinese handwriting verification algorithm using deep neural networks. In *2019 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, pages 1–2. IEEE, 2019.
- [6] Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7):971–987, 2002.
- [7] Gao Pengcheng, Gu Gang, Wu Jiangqin, and Wei Baogang. Chinese calligraphic style representation for recognition. *International Journal on Document Analysis and Recognition (IJDAR)*, 20:59–68, 2017.
- [8] Fouad Slimane, Slim Kanoun, Jean Hennebert, Adel M Alimi, and Rolf Ingold. A study on font-family and font-size recognition applied to arabic word images at ultra-low resolution. *Pattern Recognition Letters*, 34(2):209–218, 2013.
- [9] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.
- [10] Wei Tang, Yiwen Jiang, Neng Gao, Ji Xiang, Yijun Su, and Xiang Li. Scs: Style and content supervision network for character recognition with unseen font style. In *Neural Information Processing: 26th International Conference, ICONIP 2019, Sydney, NSW, Australia, December 12–15, 2019, Proceedings, Part V*, volume 11955 of *Lecture Notes in Computer Science*, pages 20–31. Springer, 2019.
- [11] Wei Tang, Yijun Su, Xiang Li, Daren Zha, Weiyu Jiang, Neng Gao, and Ji Xiang. Cnn-based chinese character recognition with skeleton feature. In *Neural Information Processing: 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, December 13–16, 2018, Proceedings, Part V*, volume 11305 of *Lecture Notes in Computer Science*, pages 461–472. Springer, 2018.
- [12] Dapeng Tao, Xu Lin, Lianwen Jin, and Xuelong Li. Principal component 2-d long short-term memory for font recognition



on single chinese characters. *IEEE transactions on cybernetics*, 46(3):756–765, 2015.

- [13] Yizhi Wang, Zhouhui Lian, Yingmin Tang, and Jianguo Xiao. Font recognition in natural images via transfer learning. In *MultiMedia Modeling: 24th International Conference, MMM 2018, Bangkok, Thailand, February 5-7, 2018, Proceedings, Part I 24*, pages 229–240. Springer, 2018.
- [14] Yixuan Wang and Yanyi Zong. Calligraphy font recognition algorithm based on improved densenet network. In *2023 Global Conference on Information Technologies and Communications (GCITC)*, pages 1–5. IEEE, 2023.
- [15] JY Xu, CY Lin, ZT Chen, ZR Deng, JH Pan, and H Liang. Handwritten calligraphy font recognition algorithm based on deep learning. *Comput Syst Appl*, 30(2):213–218, 2021.
- [16] Yong Zhu, Tieniu Tan, and Yunhong Wang. Font recognition based on global texture analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 23(10):1192–1200, 2001.