

课题基金编号 23RJGC26

华南师范大学学生课外科研一般课题立项

申 报 书

课题名称 基于深度学习通过文本对线稿图像进行着色

学科类别 信息技术

申请者 / 课题组名称 NY 组

所在学院、年级 软件学院 22 级

课题类别：

☐ 哲学社会科学类社会调查报告和学术论文

☐ 自然科学类学术论文

☒ 科技发明制作类作品

☐ 志愿服务与创业实践项目

华南师范大学制

2023 年 2 月

填写说明

填写前请先认真阅读《华南师范大学学生课外科研一般课题管理办法》的有关规定，并按下列要求认真如实填写，不要漏填、错填。由于填写不当所引起的不利于申请人的后果，责任自负。

- 1、课题基金编号由校团委统一填写并将在立项通知书中通知到各课题组，课题组成员以该课题研究成果发表论文或参赛时，须标注“**华南师范大学学生课外科研课题基金编号（具体编号）**”
- 2、申报书封面字体为**三号宋体，加粗**。
- 3、“**学科类别**”包括文科的十一大学科：哲学、经济、社会、法律、教育、管理、机械与控制（机械、仪器仪表、自动化控制、工程、交通、建筑等）、信息技术（计算机、电信、通讯、电子等）、能源化工（能源、材料、石油、化学、化工、生态、环保等）、生命科学（生物、农学、药学、医学、健康、卫生、食品等）、数理（数学、物理、地球与空间科学等）。
- 4、“**课题类别**”包括四大类别：**哲学社会科学类调查报告和学术论文**（哲学、经济、社会、法律、教育、管理）、**自然科学类学术论文**机械与控制（机械、仪器仪表、自动化控制、工程、交通、建筑等）、信息技术（计算机、电信、通讯、电子等）、能源化工（能源、材料、石油、化学、化工、生态、环保等）、生命科学（生物、农学、药学、医学、健康、卫生、食品等）、数理（数学、物理、地球与空间科学等）、**科技发明制作类作品**（与自然科学类相同）、**志愿服务与创业实践项目**（不分领域组别，参照省级以上志愿服务项目大赛获创新创业大赛的要求进行申报和培育）。
- 5、“**申请者 / 课题组名称**”：个人课题填申请者姓名，集体课题填写课题组名称。
“**所在学院、年级**”：如计算机学院 20 级。
- 6、“☐”选项填涂成“☒”形式
- 7、除封面外，均以**小四号楷体**填写。
- 8、“**学校评审委员会审核意见**”由学校评审委员会专家填写。
- 9、“**学院意见**”由学院课外科技创新领导小组领导填写课题立项意见，签章均为学院签章。
- 10、请根据实际情况，适当调整格式，以保持申报书的整洁美观。

华南师范大学学生课外科研一般课题立项申请表

课题情况	课题名称	基于深度学习通过文本对线稿图像进行着色
	课题类别	科技发明制作
	研究意义	<p>1 研究背景</p> <p>在人工智能技术不断发展的当下，绘画领域受到了巨大的冲击，出现了各种各样由人工智能辅助创作，甚至是直接生成的作品。在绘画行业中，线稿上色是一个十分消耗时间与精力的过程，如果能借助深度学习来帮助创作者去进行线稿上色，将会大大增加绘画的效率。本课题研究旨在通过基于语言的线稿自动上色技术，辅助绘画者进行填色，其研究意义不仅在插画领域、2D 动画领域，还能应用于平面设计领域。</p> <p>1.1 插画领域</p> <p>线条着色在插画行业一直是一项耗时、繁琐和精细的任务，涉及填充线稿固有色、添加阴影、细化阴影以及加入光影等步骤，画师在线稿上色之前就已经消耗大量精力于打草稿与勾线当中，在上色时又要反复的确定颜色的选择、组合，以及合适的细节设计。本课题通过画师输入线稿对应区域的颜色描述，以及一张完整的线稿，即可生成一张符合描述的上色后的图像。这将大大压缩作画所需的时间成本和精力成本，省去了填充底色、细化阴影等繁琐步骤，也节省了比较不同色彩组合的时间，使画师能够以更高的效率绘制出图片，并在生成的图像上进行细化或更改，从而提高作画效率和作品质量。</p> <p>1.2 2D 动画领域</p> <p>与插画制作相同，动画的制作同样是一件繁琐的工作，甚至所花费时间更长、耗费人力物力资源更多。商业动画的制作以逐帧动画为主，其标准以“一拍三”为主流，即每三帧放一张，每秒放八张，特殊情况下也会用到“一拍二”，甚至“一拍一”，为提高作画流畅度</p>

	<p>而牺牲了时间成本。动画的制作通常分为前期制作、中期制作、后期制作，本研究主要应用于中期制作中的动画上色部分，在原画线稿的基础上，画师通过文字描述即可得到上色后的图片，与传统的人工手动上色相比，使用线稿上色技术能缩减时间成本，特别是在需要上色数千张图像时，节约的时间成本更加显著。通过节约作画时间，画师能将时间和精力集中在更加困难的步骤，从而提升整部动画的质量，达到事半功半的效果。</p> <p>1.3 平面设计领域</p> <p>线稿自动上色在平面设计领域同样拥有着广泛的作用，能够大大提升设计师的工作效率与作品质量。线稿自动上色技术可以根据设计师的设定，自动为线稿上色，确保色彩的准确性和一致性。这可以避免因手工上色而产生的色彩不均匀或不准确的问题，从而提高设计作品的质量，使其更加专业和吸引人。除此之外，设计师在设计过程中可以以更短的时间尝试更多的色彩选择和组合，为设计师提供更多的创作思路。通过迅速获得多种颜色组合的图片，设计师能够创作出不同版本的设计图，更加高效的制作出满足甲方要求的设计。</p> <p>2 研究意义</p> <p>线稿着色若能实现自动化，将会使绘画更加经济、更加省时、更加简易。目前的图像上色方法研究可分为两大类：全自动上色与用户引导上色，随着用户自定义需求的增加，又衍生出了各种各样的着色方法。</p> <p>在众多的着色方法中，有四种根据不同的用户交互方式确立的方法：自动着色 (Automatic colorization)、基于涂鸦的着色 (Scribble-based colorization)、基于实例的着色 (Example-based colorization) 以及基于语言的着色 (Language-based colorization)，虽然这四种方法日趋成熟，但仍然有可改进之处。</p> <p>我们认为，Kumar M 等人^[23]和 Wan S^[5]等人的自动着色方法由于不具有用户的互动，虽然可以一定程度上提高着色的效率与精度，但</p>
--	--

	<p>缺少了十分重要的自定义功能,使其只能运用于现有物体的着色,而不允许稀有、不同寻常的颜色出现,在一些特定场景下无法满足用户的个性化需求。Wang T C^[2]和Yuan M^[3]等人的基于涂鸦的着色解决了上述的问题,但由于此方法预测每个像素完全依赖于用户的输入,因此会加大用户的工作量,并没有达到预想的增加效率的程度。Style2Paints 和 Xun Huang^[4]等人的基于实例的着色是通过将颜色从参考图像转移到输入图像,由此可以看出其需要至少一张现有的参考图像,如果参考图像较多,寻找一个合适的参考图像也会消耗大量的时间。因此Weng S^[5]和Manjunatha.^[16]等人的基于语言的着色是我们认为解决着色问题的最优解,然而目前的基于语言的着色方法也面临着语境混乱以及空间不一致等等问题,仍有大量的改进空间。</p> <p>本研究提出的线稿自动上色技术在以上四种方法的基础上进行改进与优化,通过画师输入线稿对应区域的颜色描述,结合一张完整的线稿,能够生成符合描述的上色后的图像,实现了对线稿的自动上色。这种方法具有以下研究意义:</p> <ol style="list-style-type: none">1. 提高绘画领域的生产效率:传统的绘画过程中,上色通常是一项耗时的工作,而线稿自动上色技术能够大大缩减上色的时间成本,使画师能以更高的效率完成作品,提高生产效率。2. 提升作品质量:线稿自动上色技术生成的图像能够保持颜色的准确性和一致性,避免了手工上色可能出现的色彩不均匀或不准确的问题,从而提升作品的质量,使其更加专业和吸引人。3. 拓展应用领域:线稿自动上色技术不仅在插画领域、2D 动画领域具有应用潜力,还可以应用于平面设计领域,帮助设计师提高工作效率,丰富色彩选择和组合,从而创作出更多版本的设计作品,满足不同需求。 <p>总而言之,本研究的线稿自动上色技术在绘画领域具有广泛的应用潜力和研究意义。通过将线稿与语言描述结合,实现了对线稿的自动上色,提高了生产效率、作品质量和应用领域的拓展。</p> <p>3 项目创新性</p> <p>目前所研发出的基于语言的自动着色普遍会出现颜色渗透、对象识</p>
--	--

	<p>别不准确、出现伪影等问题，为了解决上述问题，我们使用了 (1) 物体颜色对应矩阵^[6]OCCM (object-color corresponding matrix); (2) 注意力转移模块^[24]ATM (attention transfer module); (3) 软门控注入模块^[5]SIM (soft-gated injection module) 来进一步优化原有的模型, 使其能控制不同颜色着色的范围空间、能更加精准的根据用户输入的指令给相应的对象或物体着色、以及改善伪影等, 让基于语言的自动着色方法能真正的运用于绘画、平面设计等领域当中。</p> <p>[1]:Kumar M, Weissenborn D, Kalchbrenner N. Colorization transformer[J]. arXiv preprint arXiv:2102.04432, 2021.</p> <p>[2]:Wang T C, Liu M Y, Zhu J Y, et al. High-resolution image synthesis and semantic manipulation with conditional gans[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8798–8807.</p> <p>[3]:Yuan M, Simo-Serra E. Line art colorization with concatenated spatial attention[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 3946–3950.</p> <p>[4]: Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In Proceedings of the IEEE international conference on computer vision, pages 1501 – 1510, 2017. 2</p> <p>[5]:Weng, S., Wu, H., Chang, Z., Tang, J., Li, S., & Shi, B. (2022). L-CoDe:Language-Based Colorization Using Color-Object Decoupled Conditions. Proceedings of the AAAI Conference on Artificial Intelligence, 36(3), 2677–2684.</p> <p>[6]:Schloss, K. B., Lessard, L., Walmsley, C. S., & Foley, K. (2018). Color inference in visual communication: the meaning of colors in recycling. Cognitive Research: Principles and Implications, 3(1).</p>
--	--

		<p>[7]:Wang W, Wei F, Dong L, et al. Minilm: Deep self-attention distillation for task-agnostic compression of pre-trained transformers[J]. Advances in Neural Information Processing Systems, 2020, 33: 5776–5788.</p> <p>[8]:Dozat T, Manning C D. Deep biaffine attention for neural dependency parsing[J]. arXiv preprint arXiv:1611.01734, 2016.</p> <p>[9]:郭燕, 吴志强, 何彦辉. 一种两阶段的漫画自动着色模型[J]. 小型微型计算机系统, 2020, 41(11):2308–2313.</p> <p>[10]:Zhang B, He M, Liao J, et al. Deep exemplar-based video colorization(CJ//Proceedings ofthe IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019:8052–8061</p> <p>[11]:梁培俊. 基于深度神经网络的漫画手绘图上色方法研究[D]. 广东工业大学, 2019. DOI:10.27029/d.cnki.ggdgu.2019.000441.</p> <p>[12]: Zagoruyko S, Komodakis N. Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer[J]. arXiv preprint arXiv:1612.03928, 2016.</p> <p>[13]:Lei C, Chen Q. Fully automatic video colorization with self-regularization anddiversity[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and PatternRecognition(CVPR). 2019:3753–3761. sis, ETH Zurich, Departement of Computer Science.</p> <p>[14]:张美玉, 刘跃辉, 侯向辉等. 基于卷积网络的灰度图像自动上色方法[J]. 计算机工程与应用, 2022, 58(07):229–236.</p> <p>[15]: Xie, Y. 2018. Language-guided image colorization. Master' s the-sis, ETH Zurich, Departement of Computer Science.</p> <p>[16]: Manjunatha, V.; Iyyer, M.; Boyd-Graber, J.; and</p>
--	--	--

	<p>Davis, L. 2018. Learning to color from language. In NAACL.</p> <p>[17]:张家琛. 基于深度学习的漫画线稿图像着色方法研究[D]. 西北大学, 2020. DOI:10. 27405/d. cnki. gxbdu. 2020. 001562.</p> <p>[18]: Thomas B S, Dogra R, Dixit B, et al. Automatic image and video colourisation using deep learning[C]//2018 International Conference on Smart City and Emerging Technology (CSCET). IEEE, 2018:1-4.</p> <p>[19]:Shi X, Chen Z, Wang H, et al. Convolutional LSTM network: A machine learning approach for precipitation nowcasting[J]. Advances in Neural Information Processing Systems, 2015, 28.</p> <p>[20]: Zhang Q, Wang B, Wen W, et al. Line art correlation matching feature transfer network for automatic animation colorization[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). 2021:3872-3881.</p> <p>[21]:杨烨. 基于深度学习的视频着色方法研究[D]. 山东大学, 2022. DOI:10. 27272/d. cnki. gshdu. 2022. 005705.</p> <p>[22]:Kumar M, Weissenborn D, Kalchbrenner N. Colorization transformer[J]. arXiv preprint arXiv:2102.04432, 2021.</p> <p>[23]:Nam S, Kim Y, Kim S J. Text-adaptive generative adversarial networks: manipulating images with natural language[J]. Advances in neural information processing systems, 2018, 31.</p> <p>[24]:Xu Y, Ashley T, Brainerd EE, Bronson RT, Meyn MS, Baltimore D. Targeted disruption of ATM leads to growth retardation, chromosomal fragmentation during meiosis, immune defects, and thymic lymphoma. Genes Dev. 1996 Oct 1;10(19):2411-22. doi: 10.1101/gad.10.19.2411. PMID: 8843194.</p>
--	---

		<p>[25]:Liu, G., Fu, Y., Xu, G., Yang, S., Li, J., Wang, G., & Zhang, L. (2021). Learning a Few-shot Embedding Model with Contrastive Learning. Proceedings of the AAAI Conference on Artificial Intelligence, 35(10), 8635–8643.</p>
	技术方案	<p>4 技术方法</p> <p>4.1 总体概述</p> <p>本项目将采用 L-CoDe^[5] (a Language-based Colorization network using color-object Decoupled conditions 基于文本描述用来给线稿上色, 包括三部分。</p> <p>第一部分是一个目标对象预测器 OCCM^[6] (object-color corresponding matrix) 用于找出输入的名词和对象之间的关联。</p> <p>第二部分是 ATM 模型^[7] (attention transfer module) 将视觉区域和名词之间的对应关系转移到区域和形容词之间的对应关系, 即找出图像区域所需的色彩, 使用 OCCM 提供颜色-物体解耦条件。</p> <p>最后一部分是 SIM 模型 (soft-gated injection module) 用于调节视觉特征, 用解耦的条件。该方法在 CIE Lab 色彩空间起效, 每个图像被分为一个亮度通道 L 和两个色度通道 a 和 b。L-CoDe 预测灰度图像的两个缺失色彩通道, 最后得出上色后的图形。</p> <p>4.2 OCCM 预测器</p> <p>以前的方法^[16] (Manjunatha et al, 2018; Xie, 2018) 将标题编码为一个单一的向量, 将名词和形容词混合在一起。名词和形容词混合在一起, 导致更强的耦合问题。为了区分名词和形容词, 我们将标题中的每个词编码为一个向量。具体来说, 我们使用 Bi-LSTM (Schuster and Paliwal 1997) 来提取标题中的每个词。来提取上下文感知的特</p>

征矩阵 $W \in \mathbb{R}^{D_w \times N}$ 。 W 的每一列代表一个词向量，其维度为 D^w 。此外，我们预测此外，我们预测 OCCM 来寻找标题中名词和形容词之间的对应关系。词之间的对应关系，这也是下面的 ATM^[24] 所要求的。受 biaffine attention in dependency parsing^[7] 的启发，我们采用两个 MLPs (Dozat and Manning 2016) f^{col} f^{obj} 来将单词向量转换为 "对象空间 "和 "颜色空间"， 分别如下：

$$H_i^{col} = f^{col}(W_i) \quad H_i^{obj} = f^{obj}(W_i)$$

其中

$$H_i^{obj} \in \mathbb{R}^{D_h \times 1} \quad H_i^{col} \in \mathbb{R}^{D_w \times 1}$$

表示第 i 个分别是 "物体空间 "和 "颜色空间 "的向量。

W_i 是 W 的第 i 个列向量。颜色和物体之间的对应关系

在 OCCM 中，颜色和物体之间的相应关系是通过以下方式预测的：

$$T_{i,j} = \sigma((H_i^{obj})^T U H_j^{col} + (H_i^{obj})^T u) \sigma$$

如果第 i 个词是一个名词，其颜色被第 j 个词描述第 j 个词， T_j^i 的值就接近于 1。否则， T_j^i 就接近于 0。 $\sigma(*)$ 是 sigmoid 函数。

$U \in \mathbb{R}^{D_h \times D_h}$ 和 $u \in \mathbb{R}^{D_h \times D_1}$ 是学习参数。

4.3 ATM 模型

为了获得颜色-物体解耦的条件，跨模态注意力^{[24][12][7]} 被广泛应用 (Xu et al. 2018)，它可以协同 该方法基于单个单词的重复，对细粒度的图像进行同步分析。对其复现，可以作为一个有用的参考。我们简单回顾一下 我们简单回顾一下跨模态注意力的关键步骤，以使我们的解释 解释自成一统： 首先，他们将单词向量 投射到视觉向量的语义空间；然后他们 然后，他们计算视觉向量（查询）和词向量（关键）的点积，以获得每个视觉区域对应于句子中所有单词的概率。最后，计算单词向量（值）的加权和，以产生与每个视觉区域对应的条件。

然而，在着色任务中，输入的视觉特征缺乏颜色信息。缺乏颜色信息，所以视觉区域倾向于对应于 与名词相关。跨模态注意力忽略了

描述不寻常和反直觉的颜色细节的形容词。直观的颜色细节，最终使结果符合 "常识"。此外，这种注意使用 相同的键和值，这也加剧了耦合性。

为了克服上述问题，我们提出了一种新的 "注意力转移" 机制，将区域和名词之间的对应关系映射为区域和形容词之间的对应关系。通过这种方式，我们可以用名词来寻找对应的区域，然后用形容词来描述它们。有人提出^[25] (Liu 等人, 2021)，使用独立的目标探测层可以使关键的焦点集中在视觉区域，并对数值进行优化，以产生一个更好的条件。因此，我们使用两个线耳投影层来转换来自 "目标空间" 和 "颜色空间" 的词向量，将他们转换成视觉空间，得到 $\hat{H}^{obj} \in \mathbb{R}^{D_v \times N}$ 和 $\hat{H}^{col} \in \mathbb{R}^{D_v \times N}$ 并将它们分别作为键和值。建议的 ATM 包含 "A-T" 两个关键步骤，具体如下。

- 步骤 A: 我们通过计算视觉图像和单词的注意力图来获得区域和名词之间的对应关系：

$$S = \tilde{V}^T \hat{H}^{obj}$$

$\tilde{V} \in \mathbb{R}^{D_v \times M}$ 是 V 的扁平化版本， $M = H \times W$ ， $S \in \mathbb{R}^{M \times N}$ 。矩阵 S 中的每个元素 $S_{i,j}$ 代表第 j 个词对应于视觉特征的第 i 个位置的得分。

- 步骤 T: 为了过滤掉不合理的配对，我们将 OCCM 中小于阈值 α 的元素设为零，并将其归一为 \bar{T} (使用 l_1 -norm)。然后我们用 \bar{T} 将区域和名词之间的对应关系转移到区域和形容词之间的对应关系上，这样就得到了 S' 。图 1 显示了 S/S' 和 OCCM 的一个可视化的例子。

$$\bar{T}_{i,j} = \frac{T_{i,j}}{\sum_{k=0}^{N-1} (T_{i,k})}, S' = S \bar{T}$$

我们用 softmax 将 S' 归一化为 \bar{S}' ，并获得解耦条件 $C \in \mathbb{R}^{D_v \times M}$ 通过公式 $C_i = \sum_{j=0}^{N-1} \bar{S}'_{i,j} \hat{H}_j^{col}$ ，其中 C_i 是 C 的第 i 列，对应于第 i 个视觉特征的位置。

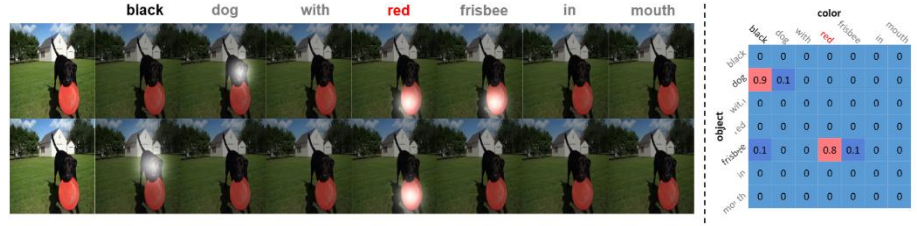


图 1^[5]

4.4 SIM 模型

当使用自然语言描述一幅图像时，人们可能只想提到他们感兴趣的对象，而其他物体则没有提供指定的颜色标签。为了预先排除这些未提及的物体在标题中出现的颜色，我们提出了 SIM 模块来处理这种颜色与物体的不匹配。

考虑到注意力模块集中在标题中提到的颜色的视觉区域（图 1 的左下角），我们考虑使用 S' 来确保颜色不会被应用于标题中没有提到的物体。具体来说，我们将 S' 重塑为 $N \times H \times W$ ，并在 S' 上应用 1×1 -conv 操作，用 sigmoid 激活 f^{conv} ，得到软门掩码 $m \in \mathbb{R}^{1 \times H \times W}$ ：

$$m = f^{conv}(S')$$

掩码 m 表示哪些视觉区域有相关的形容词。

我们采用比例和移动参数来调节视觉特征。解耦条件 C 被重塑为 $D_v \times H \times W$ ，并通过两个 1×1 -conv 层转换为尺度和移位参数。然后，我们使用软门控掩码来进一步约束调制参数：

$$\gamma' = \gamma(C) \odot m + (1 - m) \odot 1$$

$$\beta' = \beta(C) \odot m + (1 - m) \odot 0$$

元素相乘， $\gamma(*)$ 和 $\beta(*)$ 代表卷积层，分别将 C 转换为尺度和位移参数。1 和 0 分别表示 1 和 0 的张数。最后，调制后的特征 V' 可以定义为：

$$V' = \gamma' \odot \frac{V - \mu}{\sigma} + \beta'$$

μ 和 σ 是估计的平均数和标准差，由批量和空间两个维度聚合而成：

$$\mu = \frac{\sum_{b,h,w} V_{b,c,h,w}}{BHW}$$

		$\sigma = \sqrt{\frac{1}{BHW} \sum_{b,h,w} (V_{b,c,h,w}^w - \mu^2)}$ <p>其中，B 捐献的是批次大小。</p> <p>4.5 损失函数和训练</p> <p>我们采用 $\delta=1$ 的 smooth-l1 损失作为稳健的估计器来训练着色网络：</p> $\ell_{\delta}(x,y) = \frac{1}{2}(x-y)^2 \mathbb{1}_{\{ x-y <\delta\}} + \delta (x-y - \frac{1}{2} \delta) \mathbb{1}_{\{ x-y \geq\delta\}}$ <p>此外，我们对估计的 OCCM 和地面实况矩阵之间的二元交叉熵进行优化：</p> $\ell_{BCE}(x,y) = -(y \log(x) + (1-y) \log(1-x))$ <p>该模型是以端到端的方式进行训练的。在投放时，我们将灰度图像的大小调整为 $1 \times 224 \times 224$，并将其重复为 $3 \times 224 \times 224$。在编码阶段，前 4 个卷积块中包含 2（或 3）个卷积层的虚拟特征图逐渐减半，同时将特征通道的数量增加一倍。在去编码阶段，调制后的特征图在空间上逐渐增加一倍，同时将特征通道数减半。</p> <p>所有的下采样层都使用步长为 2 的 MaxPool，而所有的上采样层都使用步长为 2 的解卷积。在每个卷积块之后都添加了 BatchNorm 层，所有的卷积层都使用 ReLU 作为激活函数，只有最后一层使用 Tanht 约束输出，使其处于一个有意义的范围内。</p> <p>网络中的前四个卷积块是用 VGG16-BN 模型的预训练权重进行初始化的。着色网络的其余部分是用 Xavier 方法初始化的。</p> <p>在 ATM 中，我们将批次大小设置为 16，$\alpha=0.1$。</p> <p>我们使用亚当优化器使我们的目标损失最小化，学习率设定为 2×10^{-4}，动量参数 $\beta_1=0.99$ 和 $\beta_2=0.999$。</p>
--	--	---

5 技术可行性

L-CoDe 在灰度图像上色^[6]表现的效果显著，因此本项目将采用 L-CoDe 基于文本描述用来给线稿上色，其中 OCCM 和 ATM 用于解决颜色与对象耦合问题，SIM 用于来处理颜色-对象不对应导致的不匹配问题。并且提出一个新的数据库，包含颜色-对象对，以提供解决耦合问题的相关信息。

5.1 L-CoDe 可行性



图 2：与另外三种方法的定性比较^[5]

Table 1: Quantitative comparison result. L-CoDe (ours) performs best in three metrics. Throughout this paper, \uparrow (\downarrow) means higher (lower) is better.

Method	PSNR \uparrow	SSIM% \uparrow	LPIPS \downarrow
(Manjunatha et al. 2018)	21.055	85.333%	0.282
(Xie 2018)	21.407	84.016%	0.298
Ours	24.965	91.657%	0.169

Table 2: User study result. We conduct two user study experiments to evaluate whether our colorization results are favored by human observers. L-CoDe (ours) achieves obviously higher scores in both experiments.

Method	Experiment-1	Experiment-2
(Manjunatha et al. 2018)	18.32%	16.68%
(Xie 2018)	18.68%	32.64%
Ours	63.00%	50.68%

从 L-CoDe 的实验数据看出, L-CoDe 成功地解决颜色对象不匹配和耦合问题, 正确地修改了所需对象的颜色 (顶部), 并避免了错误地为标题中未提及的对象着色 (底部)。

在 Table 1^[5]中显示了以对应标题为条件的着色结果的定量比较。L-CoDe 在所有三个指标上都优于 Xie^[15]和 Manjunatha et al.^[16]的方法, 这表明 L-CoDe 的结果在这些指标上更类似于地面实况图像。如图所示。通过交换标题中的不同颜色来生成单个图像的不同颜色图像, 评估出每种方法是否能够精确地为对应的对象着色。这些例子表明, L-CoDe 在处理颜色对象耦合和颜色对象不匹配方面特别有效。

L-CoDe 的开发人员在 AMT (Amazon Mechanical Turk) 上发布的两次实验^[5]中, 实验一提供了描述地面实况彩色图像的字幕, 向参与者展示了地面实况图像和来自三种不同方法的三种彩色结果, 并要求他们在考虑地面实况的情况下选择视觉上最佳的结果。而实验二中, 实验者用另一个随机的颜色词替换标题中的颜色词, 参与者会看到一个新的标题和三种不同方法的三种彩色结果, 并被要求选择一张与给定标题最匹配的图像。每个实验由 100 个任务组成, 每个任务中的图像是从测试数据集中随机选择的, 且每个任务由 25 名参与者完成。最终结果如 Table 2 所示, L-

CoDe 在实验中取得了更高的分数,这也验证了 L-CoDe 的可行性。

5.2 ATM 和 SIM 在消融实验中的有效性^[10]

Table 3: Quantitative comparisons with replacing ATM (1), replacing SIM (2) and our complete method (Ours).

Method	PSNR↑	SSIM↑	LPIPS↓
(1)	24.778	91.643%	0.175
(2)	23.497	89.869%	0.207
Ours	24.965	91.657%	0.169

在更换掉 SIM 模块后生成的彩色图像中存在色散和失配。由表 3^[5]得出,完整地采用 ATM 和 SIM 模块在所有三个指标上都获得了更高的分数。

6 课题扩展可行性

在本项目基础研究完成后,有着足够且广阔的拓展空间,本项目后期可拓展研究:

1. 二次着色^[9]:针对用户语言引导的辅助着色方法生成的彩色化结果,可能会出现一些小区域未着色或与预期结果有偏差的问题。为了改善这些问题,可以采用二次着色方法对不理想的区域进行颜色纠正,从而进一步优化着色效果。这样,在不需要专业着色技巧的情况下,也能获得更好的彩色化结果。



2. 将线稿图像自动着色应用到视频中^[9]:本研究的主要工作是对黑白线稿图像进行着色,而尚未运用到视频中。对于连续图像组成的视频,可对本研究方法进行深入拓展,将线稿图像自动着色应用到视频和实践工作中。

3. 基于深度学习的视频着色方法^[13]:可分为两类,自动视频着色方法和基于参考的视频着色方法。自动视频着色方法不需要额外加入提示信息,网络经过大规模的数据集训练之后,输入待着色视频帧就能自动为视频着色。Thomas^[10]在卷积网络中加入长短时记忆网络(Long Short Term Memory Network, LSTM)^[11]来为图像或视频着色。该方法通过卷积网络提取输入灰度视频帧的特征,同时采用 LSTM 利用已着色的前一视频帧来指导当

		<p>前帧的着色，从而加强视频帧之间的联系。Lei^[13]提出带有自正则化且能为视频多样化着色的自动视频着色方法。该方法使用多样化损失产生不同的着色结果，使用自正则化加强相似颜色像素之间的信息传递，保持视频中时空颜色的一致性。基于参考的视频着色方法需要用户提供一个或者多个彩色图像作为参考图像指导着色，通过网络将参考图像中的颜色迁移到待着色视频中去。Zhang^[12]提出为线稿动画着色的框架。框架中加入了可学习的相关性匹配特征迁移模型 (Correlation Matching Feature Transfer model, CMFT) 来匹配参考特征与线稿特征之间的相关性，并以特征尺寸逐渐递减的方式将模型集成到基于 U-Net 结构的生成器中。</p> <p>4. 将 L-CoDe 模型运用于漫画自动着色^{[9] [11]}中: 由于漫画的原始线稿中并没有漫画人物的纹理细节、阴影等信息，所以相比现有的照片着色生成任务，漫画着色的难度大大增加，而用本研究所引用的 L-CoDe 模型进行线稿着色可以大大提高漫画着色效率。</p>
课题情况	研究计划	<p>开始时间：2023 年 4 月</p> <p>2023 年 4 月-9 月</p> <p>完成基础模型的研究，对比不同现有模型改进。</p> <p>2023 年 9 月-12 月</p> <p>训练出可以使用的模型，并基于线稿加强</p> <p>2024 年 1 月- 3 月</p> <p>进行相关实验，推广产品，进行封测</p> <p>2024 年 4 月</p> <p>上线项目，申请软著, 完结项目</p> <p>完成时间：2024 年 4 月</p>

	预期成果	完整代码一份 软件著作权一份 申请专利一份					
	作品展示形式	<input type="checkbox"/> 实物、产品 <input checked="" type="checkbox"/> 模型 <input type="checkbox"/> 图纸 <input type="checkbox"/> 磁盘 <input checked="" type="checkbox"/> 现场演示 <input type="checkbox"/> 图片 <input type="checkbox"/> 录像 <input type="checkbox"/> 样品					
	经费预算和领取方式	总 计： 2000 （元人民币） 具体预算： 实验材料费 500 元 论文发表费 500 元 专利申请费 1000 元					
	其他说明						

课 题 组 成 员	负 责 人	姓名	性别	学院	学历	联系电话	其他通讯
		吴嘉桐	男	软件学院	本科	13509663771	20222005123@m.scnu.edu.cn
	课 题 参 加 人	周驰恒	男	软件学院	本科	18603039376	3202329956@qq.com
		殷嘉豪	男	软件学院	本科	15019263270	1228749586@qq.com

指导教师	姓名	职称	学院	联系电话	其他通讯
	张承忠	讲师	软件学院	13927261183	zcz505@163.com
	<p>指导教师推荐意见：该项目利用深度学习技术通过文本对线稿图像进行着色，推荐申报。</p> <p style="text-align: right;">张承忠 签名：</p>				
学院 评审 委员 审核意 见	<p style="text-align: center;">同意推荐申报</p> <p style="text-align: right;">  签名（章）： </p>				
学院 意见	<p style="text-align: center;">同意</p> <p style="text-align: right;">  签名（章）： </p>				
学校评 审委员 审核意 见	<p style="text-align: right;">签名（章）：</p>				