



基于局部感知的文字图像合成

连宙辉

王选计算机研究所



北京大学
PEKING UNIVERSITY

研究背景

文字是人类用表义符号记录表达信息以传之久远的方式和工具。人类长期只有口语传递信息，用文字记事传递信息形成的**书面语历史较短**。

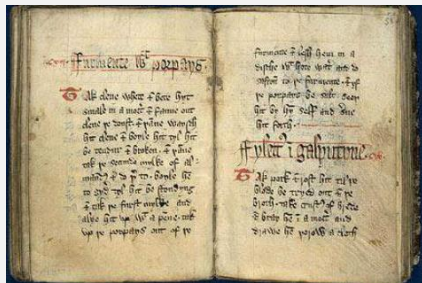
象形文字(hieroglyph)

由图画文字演化而来的，
是一种最古老的字体
埃及的象形文字、苏美尔文、古印度文以及**甲骨文**



表音文字(Phonogram)

使用少量的字母记录语言中的语音，记录语言的文字。
英语、法语、俄语、阿拉伯语、维吾尔语、日文假名、蒙古语、韩语、德语等



意音文字(logogram)

一种图形符号既代表语素，又代表音节的文字系统。
汉字、楔形文字、圣书字、契丹文、女真文/满文、西夏文、越南字喃



研究背景

文字使人类进入有历史记录文明社会，是人类文明的重要标志



仓颉造字



秦始皇书同文



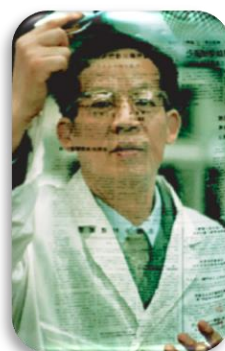
蔡伦造纸术



书圣王羲之



毕昇印刷术



王选汉字激光照排



“汉字是中华文明的重要标志，也是传承中华文明的重要载体”——习近平主席



陶文



甲骨文



金文



简牍



雕版印刷



活字印刷



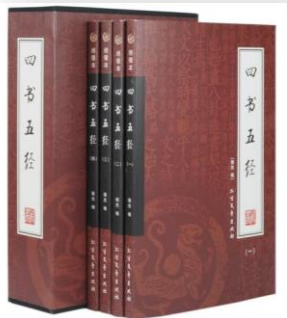
激光照排



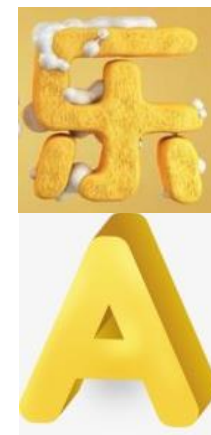
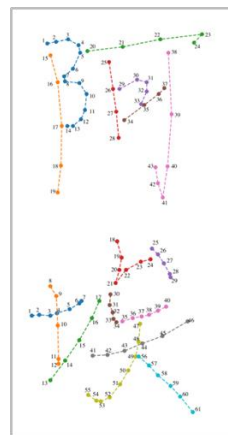
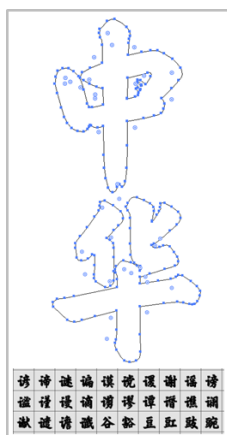
智能终端

研究背景

文字应用无处不在



字形表现形式多样



亟待形状建模与图像合成技术的研究与突破

研究背景

形状建模与图像合成

方法

复用编辑

数据生成

全局感知
建模

局部感知
合成

风格迁移
生成

核心挑战

真实性

小样本

创造力

科学问题

精准学习

类人推理

鲁棒泛化

研究背景

形状建模与图像合成



聚焦
文字应用



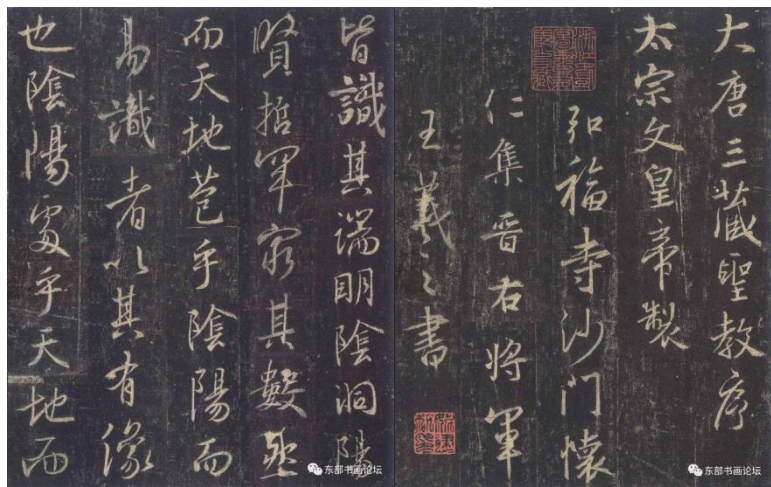
“让机器具备人类水准的**图形图像设计与创作能力**”
解决现有AI“有多少人工，就有多少智能”的痛点

研究背景

案例1

中共中央办公厅、国务院办公厅2017年发布
《关于实施**中华优秀传统文化传承发展工程**的意见》

“中华精品字库工程”是中华优秀传统文化传承发展工程支持项目
工程将精选**100位中国历代书法名家**作品,开发成电脑字库
弘扬优秀传统文化,传承中华汉字之美。



现有
工艺

收集
字稿

字稿
扫描

字形
选择

精细
设计

补字
创写

字形
检查

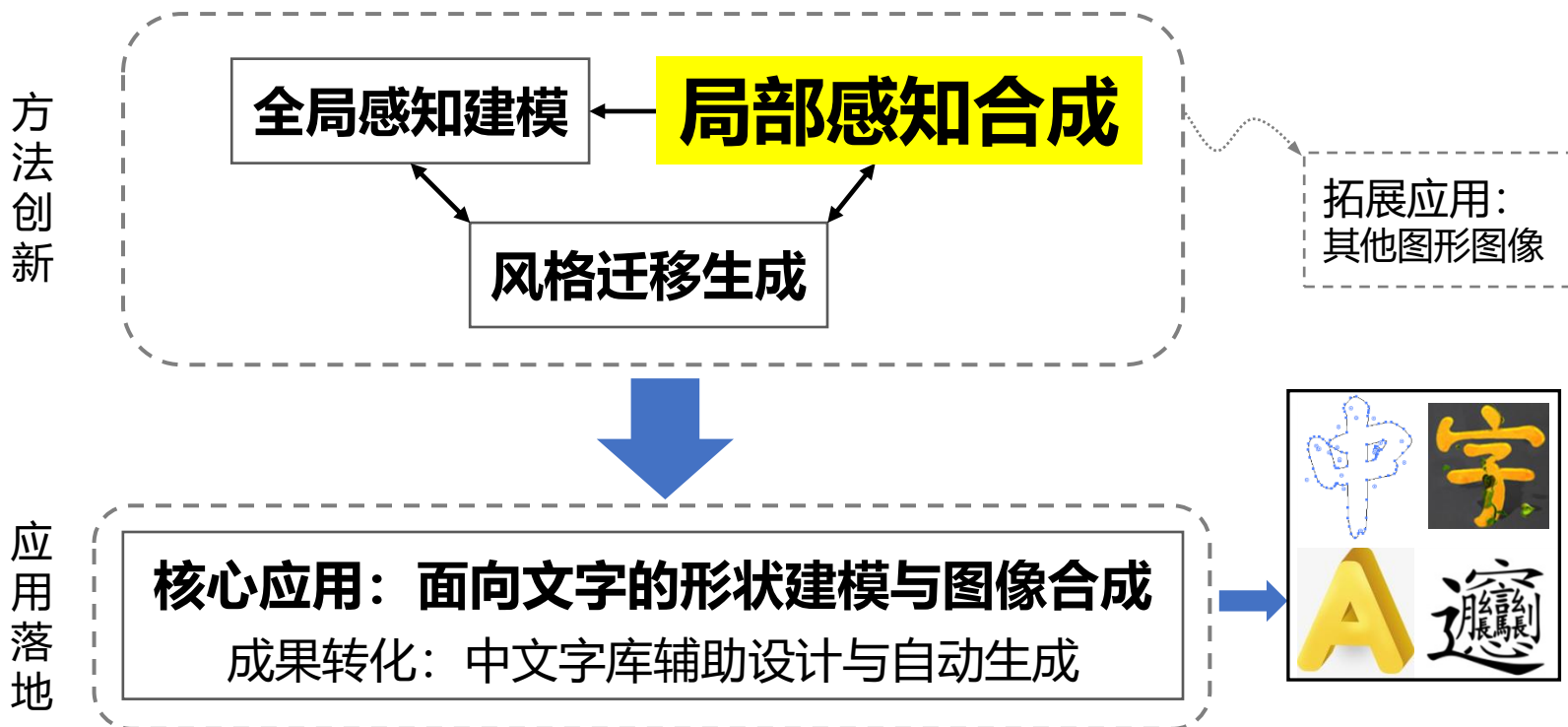
专家
审核

封装
成库

亟待面向文字的**形状建模与图像合成技术**的研究和突破

研究背景

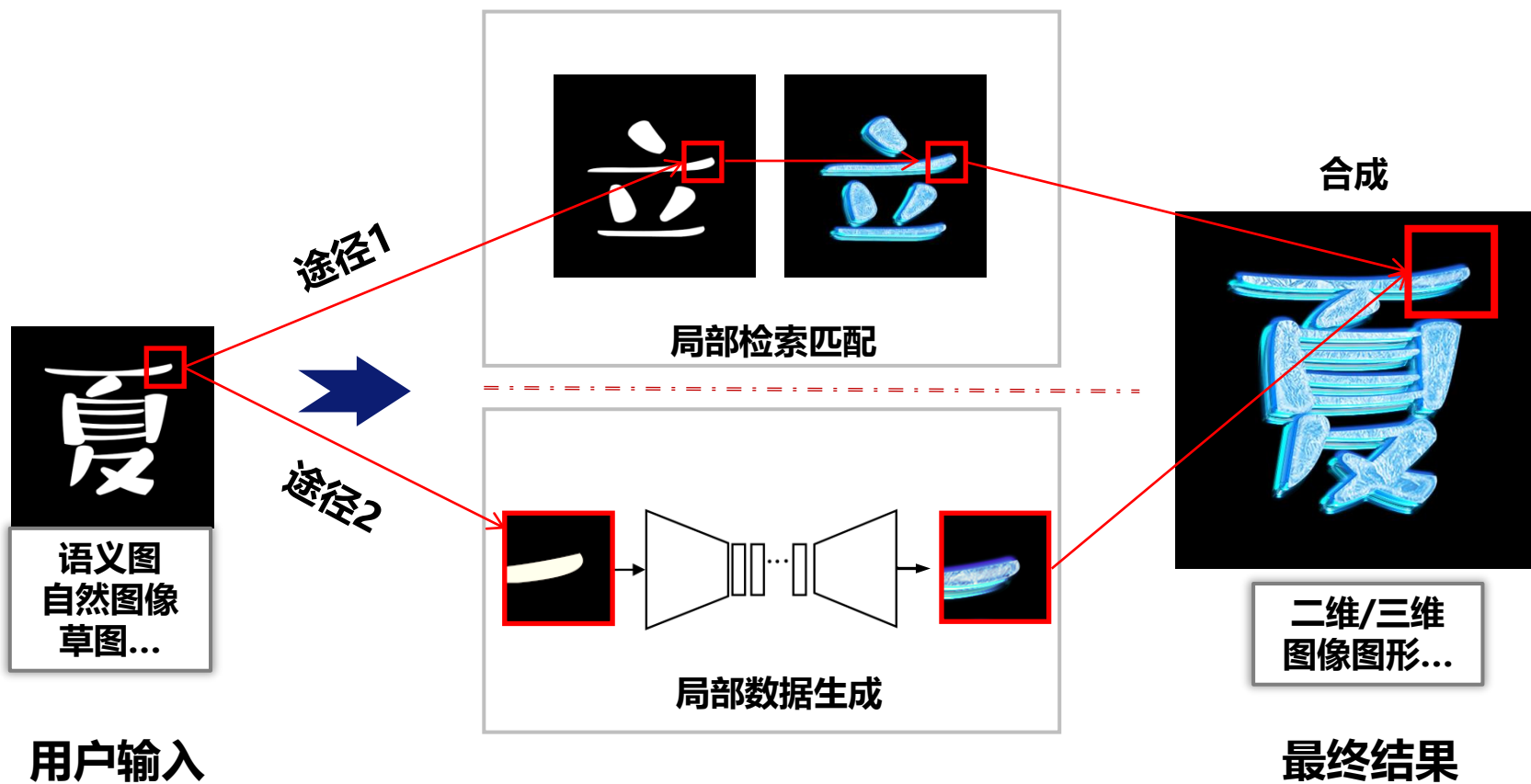
2011年至今聚焦面向文字的形状建模与图像合成技术研究



研究进展

局部感知合成

采用局部特征检索匹配，经由局部复用生成实现图形图像生成



主要
挑战

局部精准描述、高效匹配检索、保真复用生成

研究进展(1/3)

A Common Framework for Interactive Texture Transfer

Y. Men, Z. Lian*, Y. Tang, J. Xiao

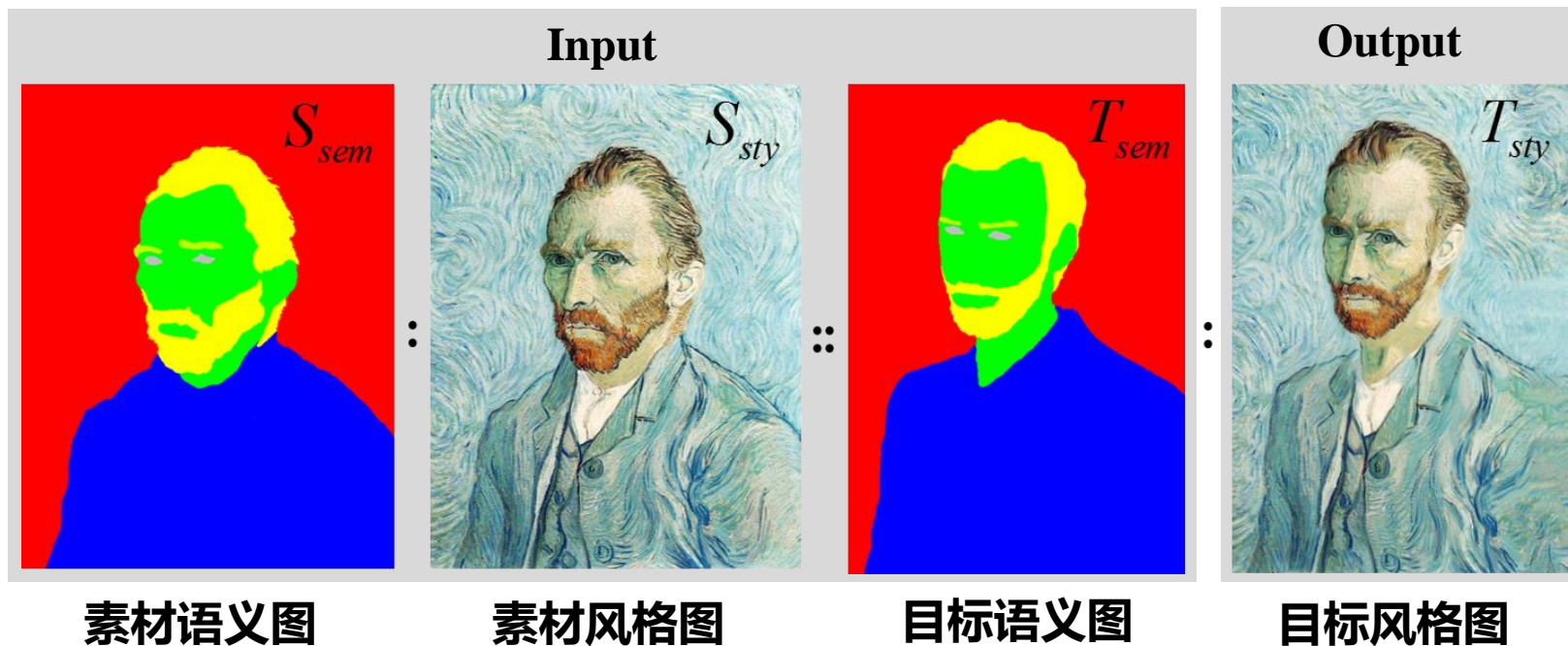
CVPR 2018 (Spotlight)



北京大學
PEKING UNIVERSITY

研究进展(1/3)

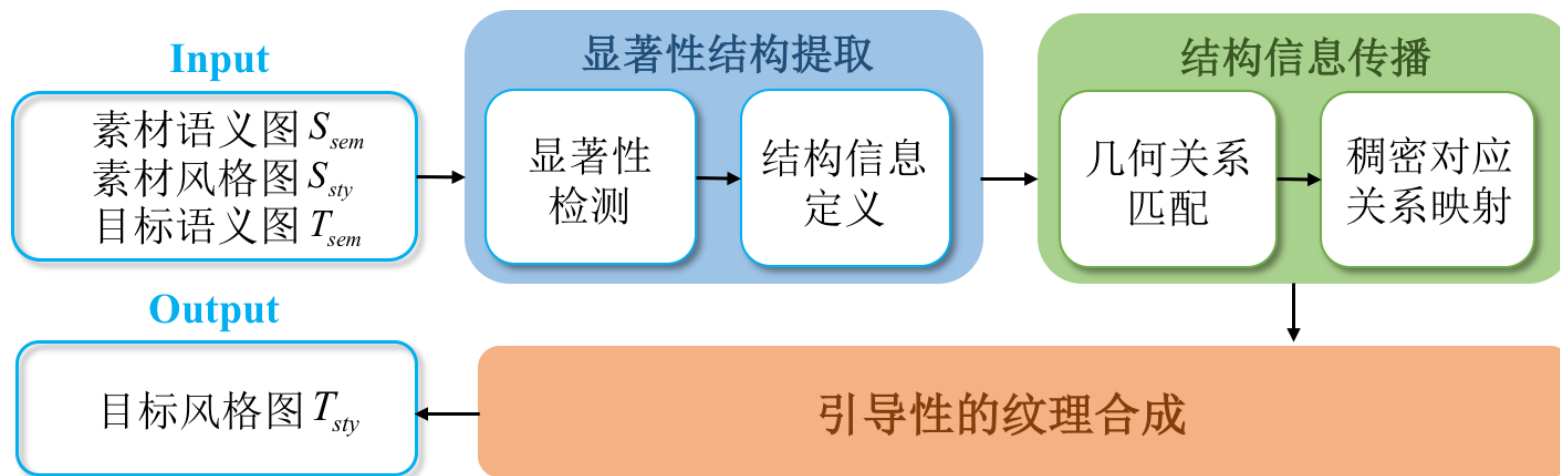
- 问题描述



研究进展(1/3)

• 解决方案

融合结构先验的纹理迁移框架



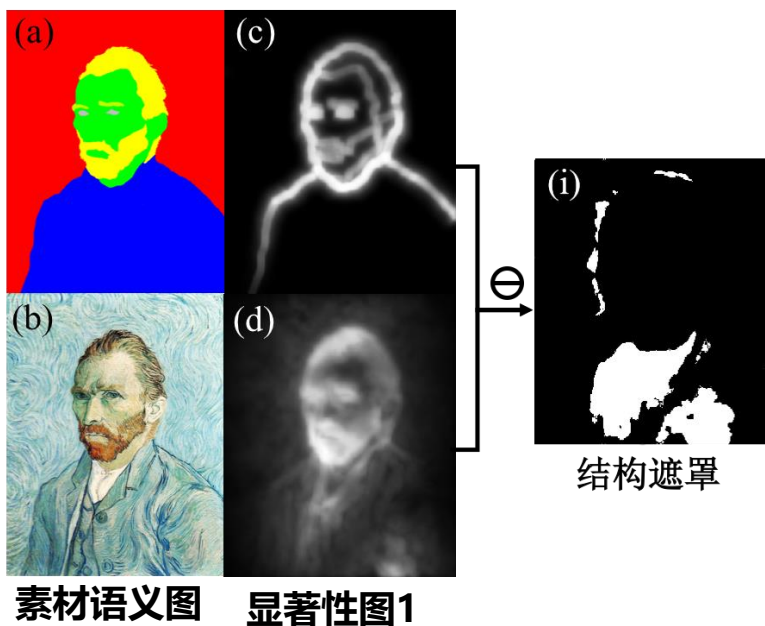
研究进展(1/3)

• 解决方案

(1) 显著性结构提取

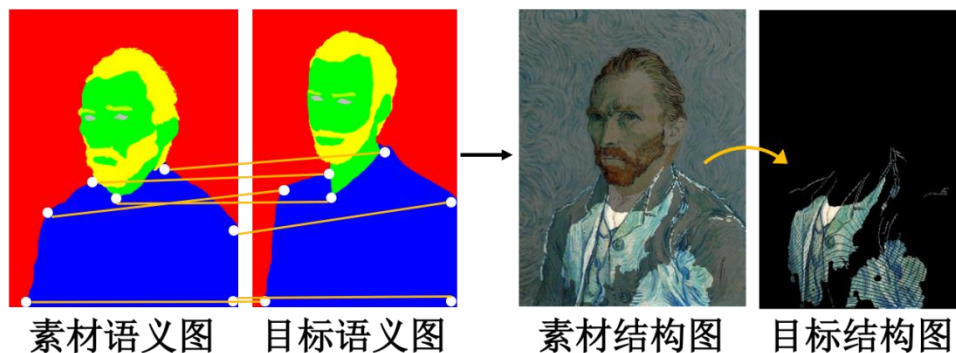
(2) 结构信息映射

素材风格图 显著性图2



轮廓关键点匹配 (CPD)

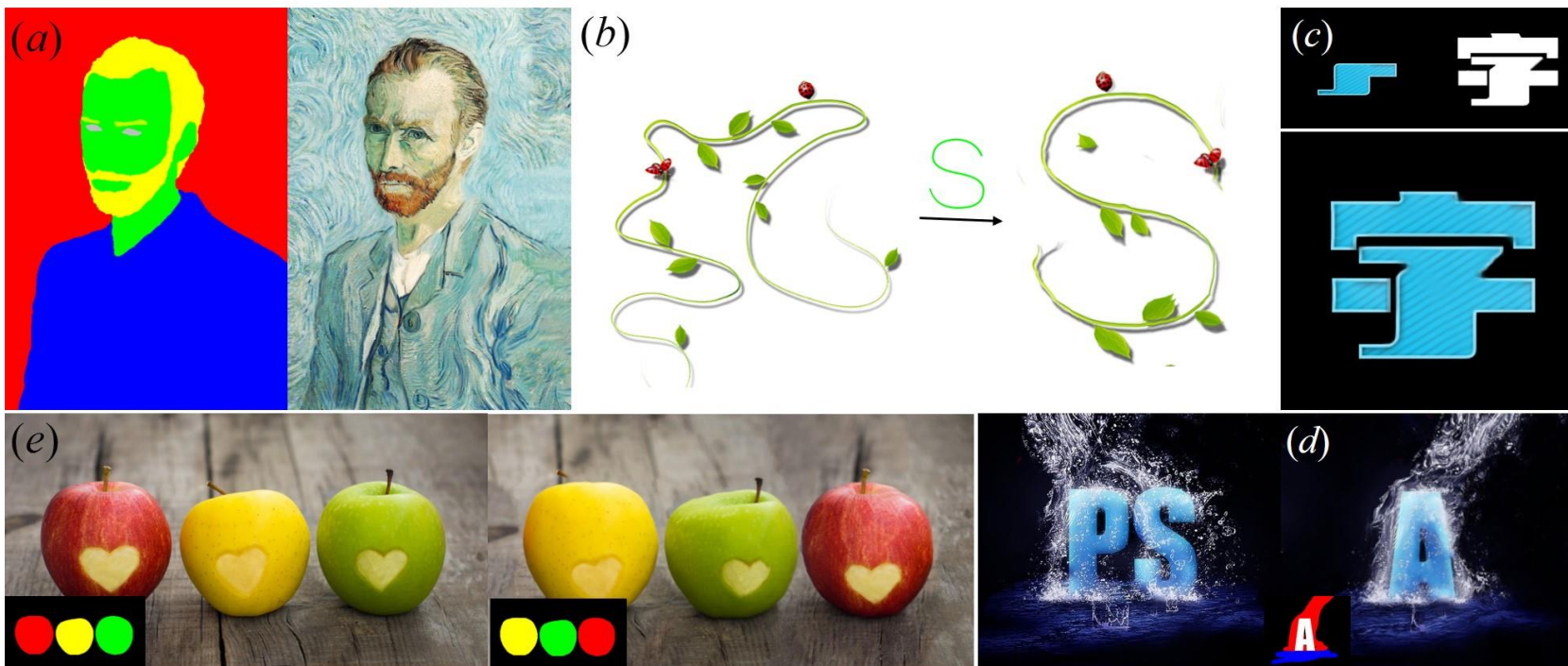
薄板样条插值 (TPS)



研究进展(1/3)

• 多任务场景下生成高质量纹理迁移结果

子任务: (a)涂鸦变画作 (b) 装饰性纹理路径编辑
(c,d)特效字生成, 控制特效纹理的空间分布 (e)纹理交换



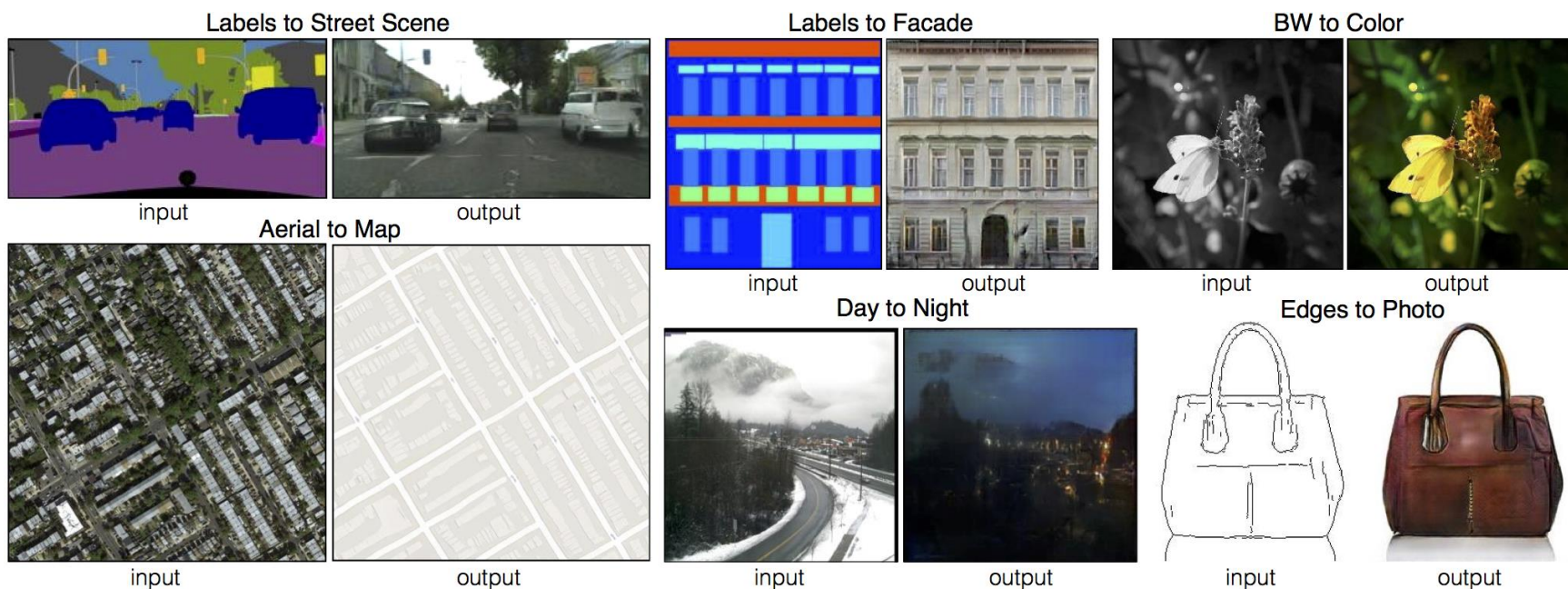
基于传统纹理合成模型实现

研究进展(1/3)

• 相比于深度生成模型 (pix2pix, pix2pixHD, cycleGAN,...)

(1) 深度生成模型依赖于**单一类别下大规模的训练数据**，只能用于**有限的应用场景**中，如街景图、建筑物、人脸的生成。

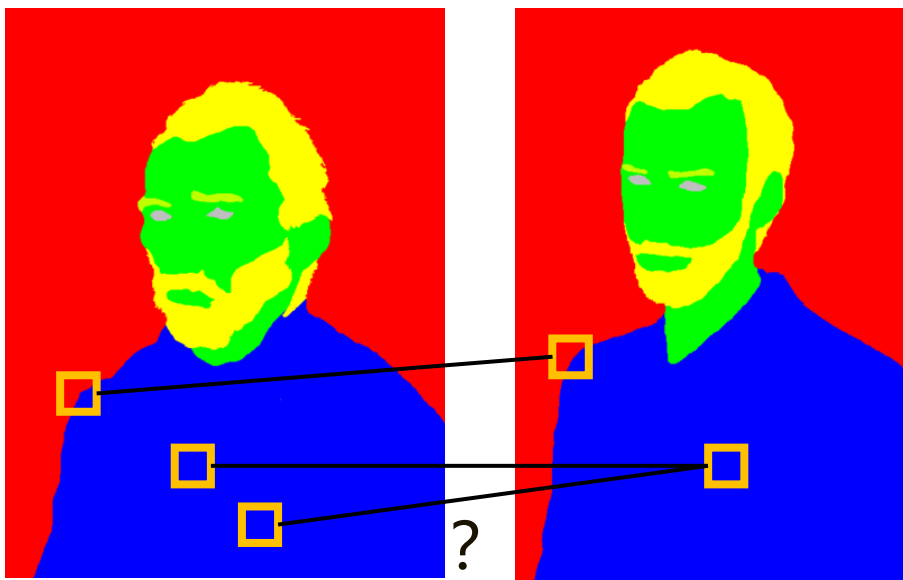
(2) 实际应用中无法获取任意一张图像在不同角度、布局下的成对训练数据。



研究进展(1/3)

• 核心思想

- 纹理合成 -> 能量优化问题
- 多个自定义引导项构建目标方程，动态引导合成过程



- 色块边缘
语义信息丰富、包含更多特征
语义引导项

- 色块内部
依赖纹理连续性 (边缘->内部) 传播
易造成内部纹理错乱
纹理相干项、结构引导项

降解为图像修复问题：破损填充，从边缘向内部进行纹理拓展

研究进展(1/3)

• 实现方法

引导性的纹理合成

目标方程 $E = \sum_{q \in T} \min_{p \in S} (\lambda_1 E_{semantic}(p, q) + \lambda_2 E_{texture}(p, q) + E_{structure}(q))$

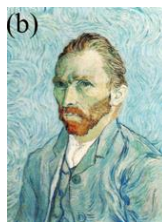
(1) 语义引导项

$$E_{sem}(p, q) = \|T_{sem}(N_q) - S_{sem}(f(N_p))\|^2 \quad f(N_p^i) = \gamma H \Delta N_p^i + p \quad H = \begin{vmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{vmatrix}$$



(2) 纹理相干项

$$E_{corhence}(p, q) = \|T_{sty}(N_q) - S_{sty}(f(N_p))\|^2$$

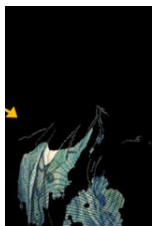


(3) 结构引导项

$$E_{struct}(q) = \sum_{i=0 \dots w^2-1} \frac{(T_{struct}(N_q^i) - T_{sty}(N_q^i))^2 \lambda(N_q^i)}{\tau(N_q)}$$

$$\lambda(N_q^i) = \begin{cases} 1, & \lambda(N_q^i) \in \Omega'_{struct} \\ 0, & \lambda(N_q^i) \notin \Omega'_{struct} \end{cases}$$

$$\tau(N_q) = \sum_{i=0 \dots w^2-1} \lambda(N_q^i)$$



研究进展(1/3)

• 实现方法

引导性的纹理迁移

$$\text{目标方程 } E = \sum_{q \in T} \min_{p \in S} (\lambda_1 E_{\text{semantic}}(p, q) + \lambda_2 E_{\text{texture}}(p, q) + E_{\text{structure}}(q))$$

迭代的最大期望值算法 (EM算法)

E步: 固定目标图像和素材图像之间的匹配关系, 进行目标图的重建

→ 图像重建

M步: 固定目标图像, 查找更优的目标图像和素材图像之间的匹配关系

→ 匹配关系搜索

匹配关系 $NNF \rightleftharpoons I'$ 目标图像 $\longrightarrow E \downarrow$

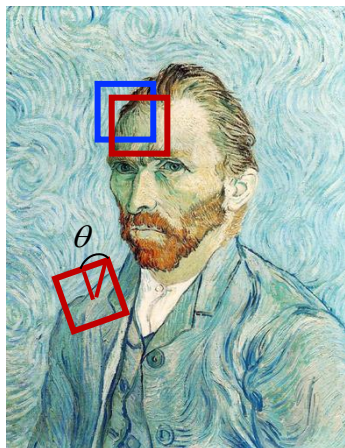
研究进展(1/3)

• 实现方法

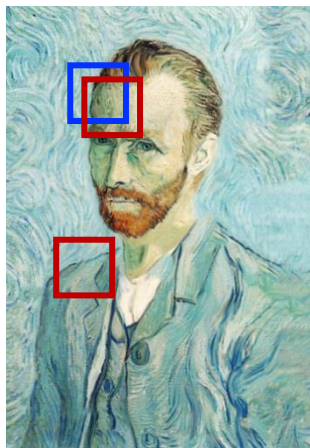
引导性的纹理迁移

- 匹配关系搜索

采用改进的PatchMatch算法



素材风格图



目标风格图

改进1: 拓展搜索空间

$$[x, y] \rightarrow [x, y, \theta, \gamma]$$

改进2: 采用矩阵运算

优点: 在更短时间内获得更丰富的素材资源

研究进展(1/3)

• 实现方法

Algorithm. 1. Guided texture transfer

Input: $S_{sem}, T_{sem}, S_{sty}$

Output: T_{sty}

1: Compute T_{struct} via structure extraction and propagation

2: Create image pyramids $S_{sem}^l, T_{sem}^l, S_{sty}^l, T_{struct}^l$

3: Guided initialize NN^{L-1}, T_{sty}^{L-1}

4: for scale $l = L - 1$ to 0 do

5: for iteration $j = 0$ to m do

6: $NN^l \leftarrow search(S_{sem}^l, T_{sem}^l, S_{sty}^l, T_{sty}^l, T_{struct}^l, NN_l)$

7: $T_{sty}^l \leftarrow vote(NN^l, S_{sty}^l)$

8: end for

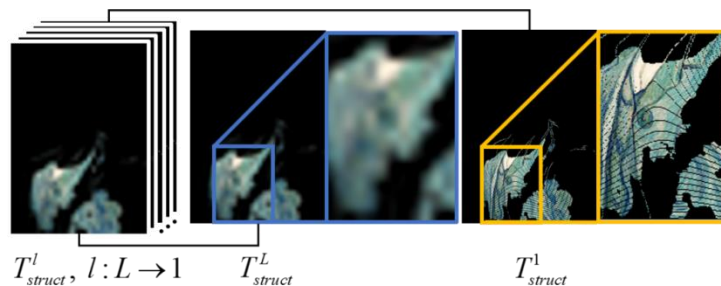
9: if $l < L - 1$ then

10: $NN^{l-1} \leftarrow Upsample(NN^l)$

11: $T_{sty}^{l-1} \leftarrow Upsample(T_{sty}^l)$

12: end if

13: end for



EM
iteration

Initialize for
next level

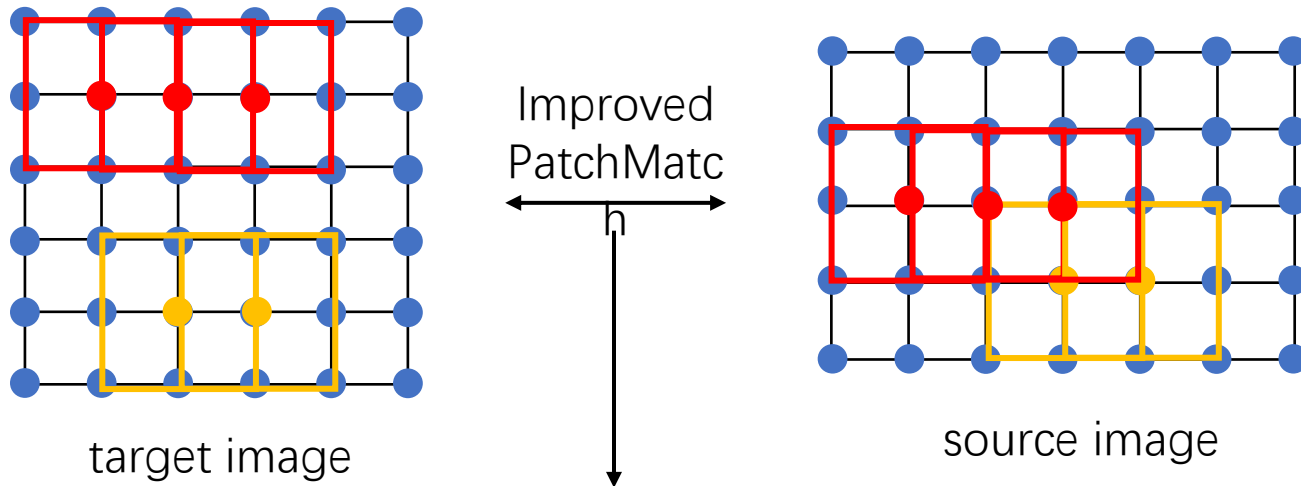
Image pyramid
(coarsest-
>finest)

研究进展(1/3)

• 实现方法

Search — find nearest-neighbor field

Propagation [right, down, left, up]
Random search



NNF(nearest neighbor field)

$[x, y, \theta, \gamma]$

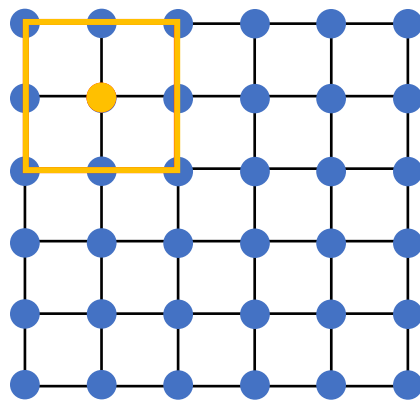
研究进展(1/3)

• 实现方法

Search — find nearest-neighbor field

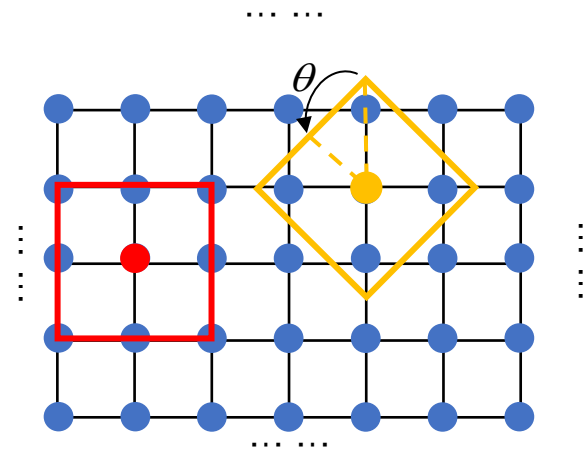
Propagation [right, down, left, up]

Random search



target image

Improved
PatchMatc
← h →



source image

NNF(nearest neighbor field)

$[x, y, \theta, \gamma]$

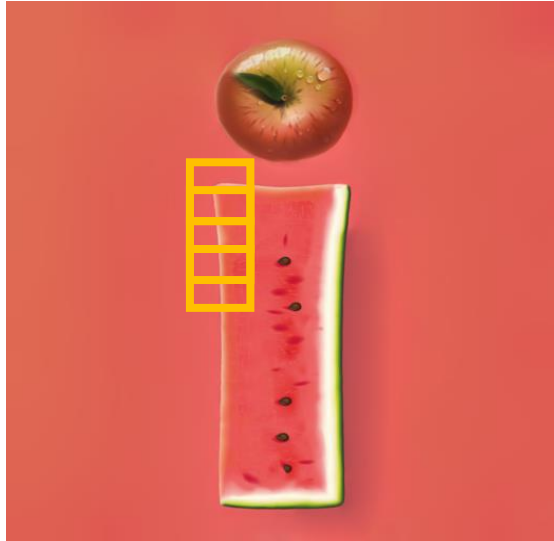
研究进展(1/3)

• 实现方法

Search — find nearest-neighbor field

Propagation [right, down, left, up]

Random search

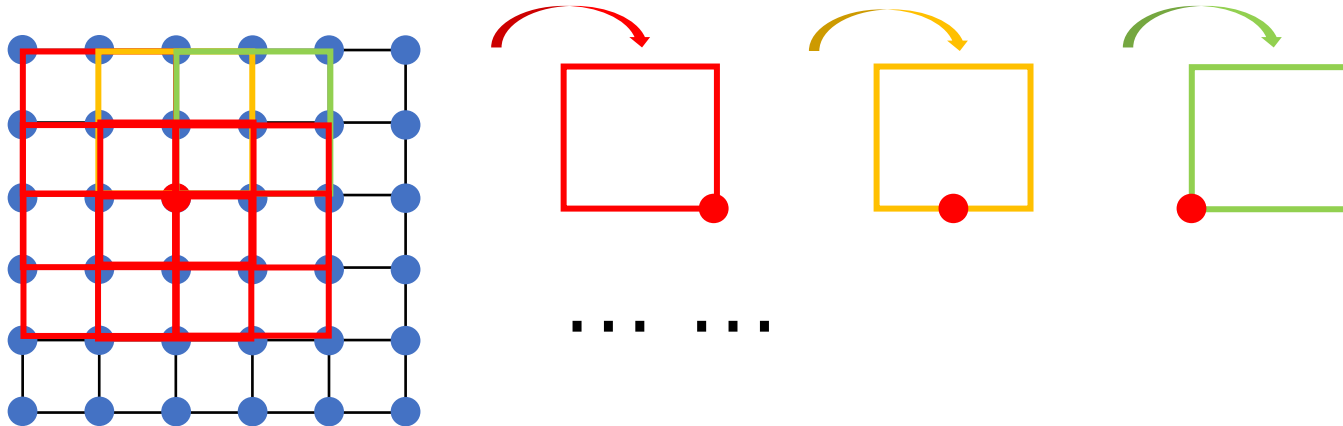


propagation in the direction of rotation

研究进展(1/3)

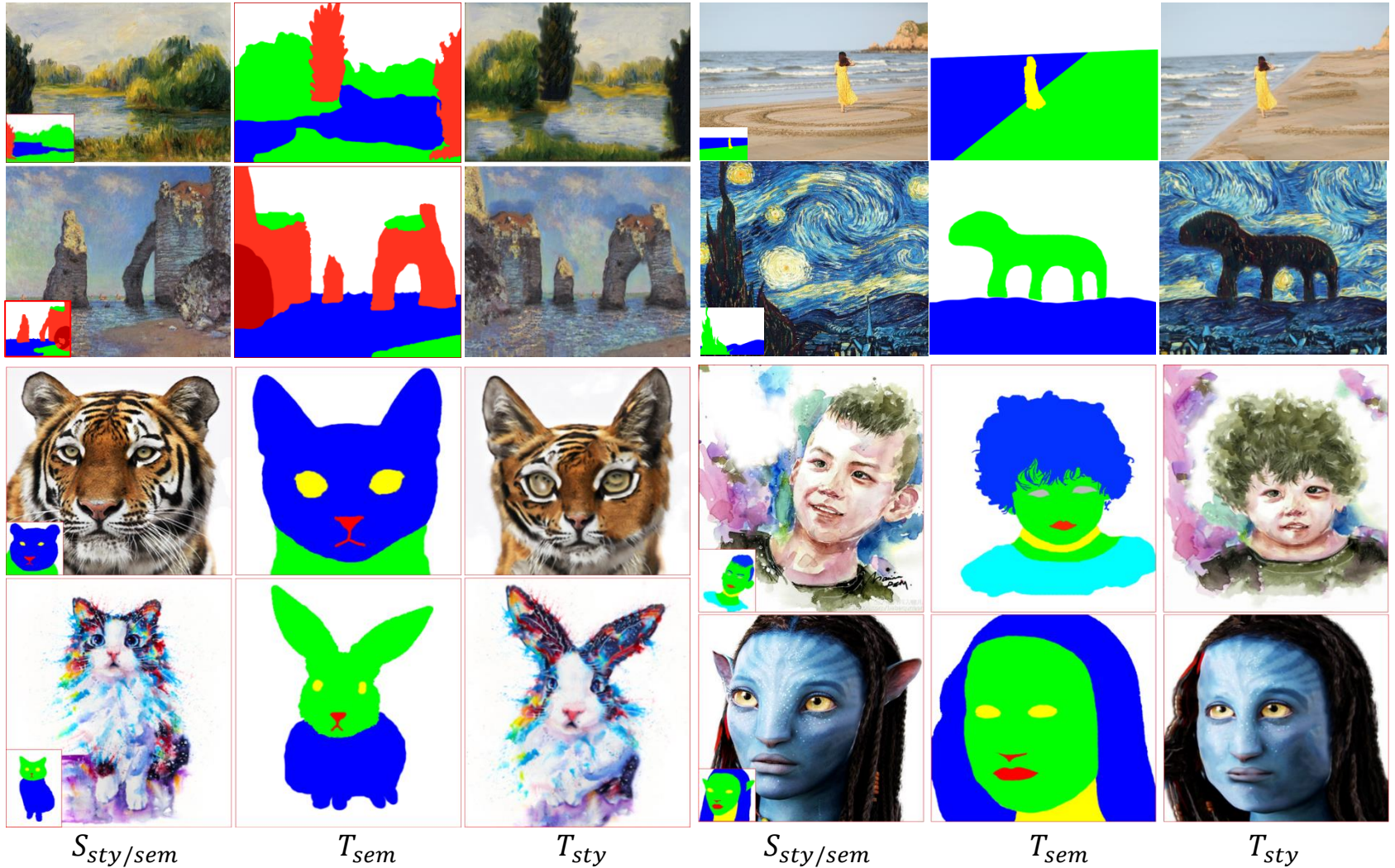
- 实现方法

Vote — reconstruct target image



研究进展(1/3)

• 实验结果 涂鸦变艺术画作



研究进展(1/3)

• 实验结果 装饰性纹理路径编辑



Input (path 1)



Input (path 2)



Input (path 1)



Input (path 2)



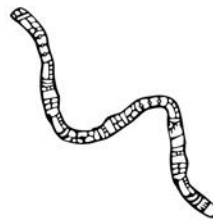
Input (source)



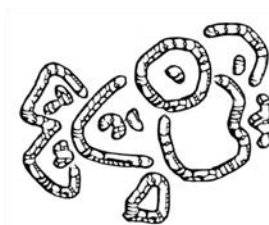
Output 1



Output 2



Input (source)



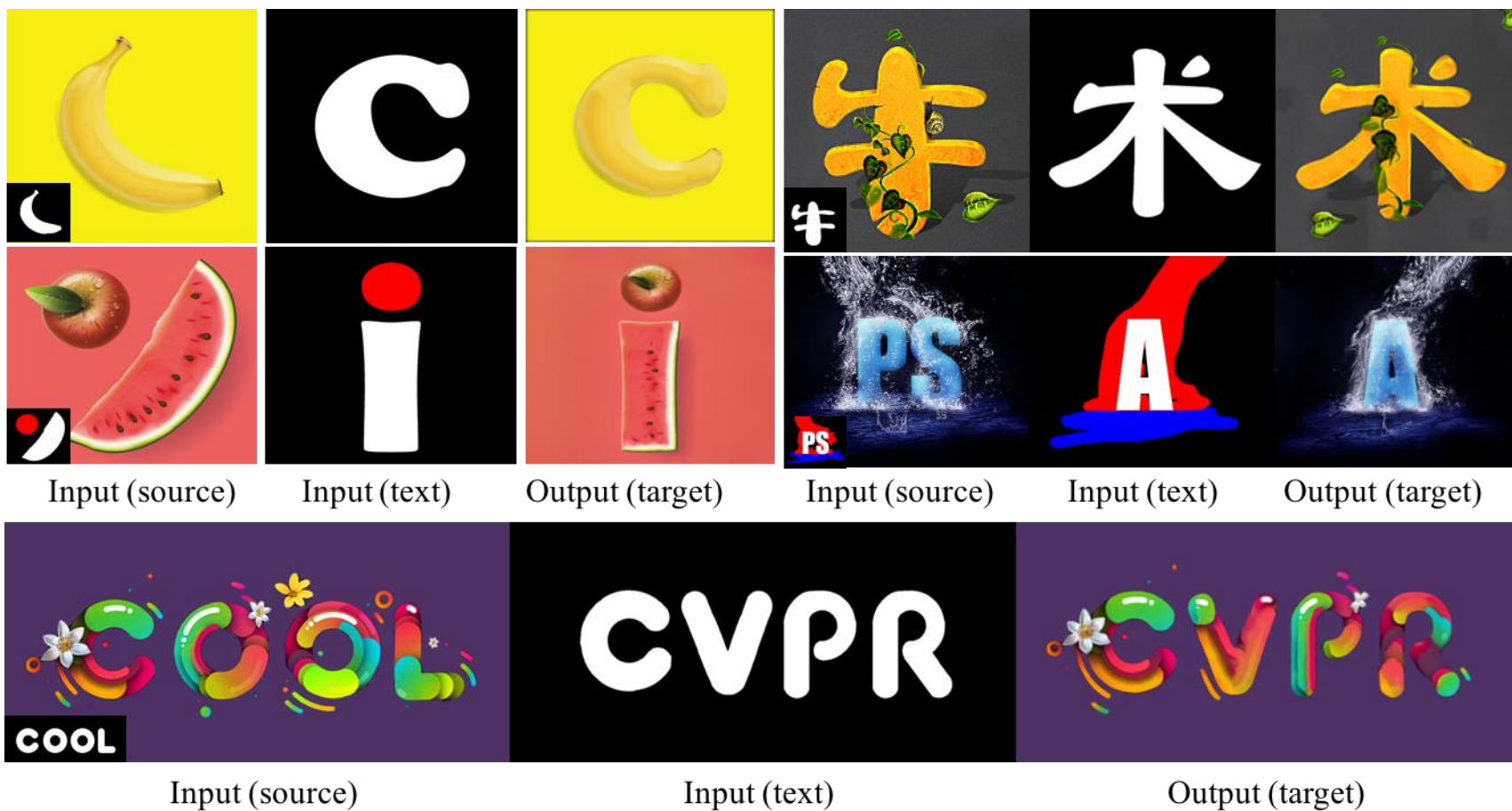
Output 1



Output 2

研究进展(1/3)

• 实验结果 特效文字生成



研究进展(2/3)

DynTypo: Example-based Dynamic Text Effects Transfer

Y. Men, Z. Lian*, Y. Tang, J. Xiao

CVPR 2019



北京大學
PEKING UNIVERSITY

研究进展(2/3)

- 场景介绍

Style Exemplar

研究进展(2/3)

- 问题描述

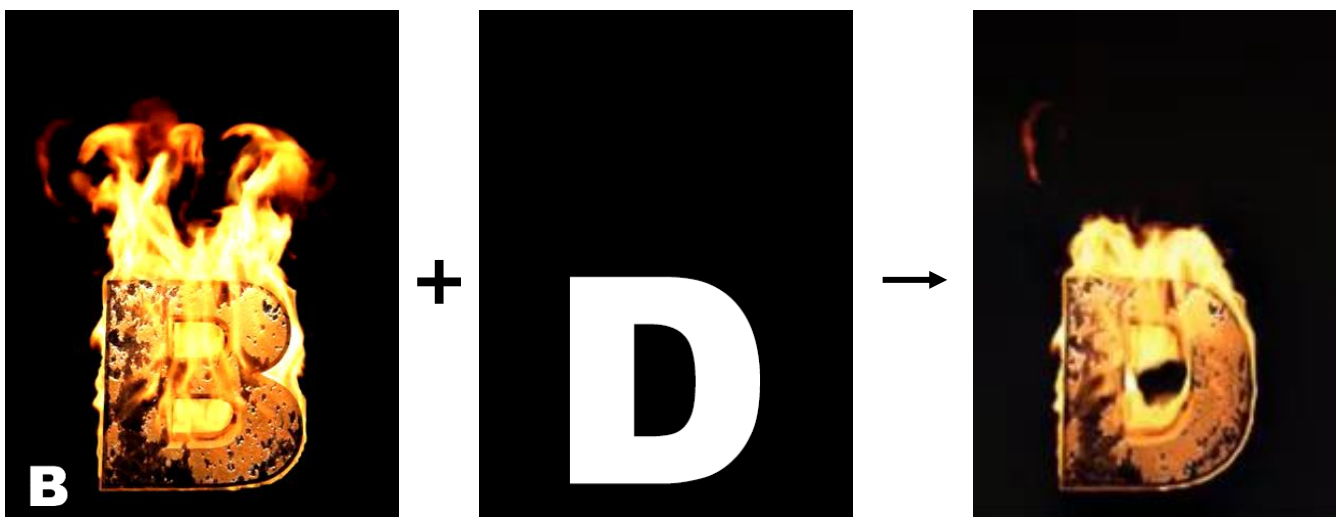
时序空间的拓展



研究进展(2/3)

• 问题分析

最简单方案：直接逐帧进行图像纹理迁移



- **时序**问题： 时序不稳定，生成结果发生明显纹理突变、闪烁
- **外观**问题： 强烈的动态纹理效果难以完全迁移

研究进展(2/3)

• 问题分析

常用优化方案：逐帧图像纹理迁移+时序限制

单张目标图像 -> 无目标运动流



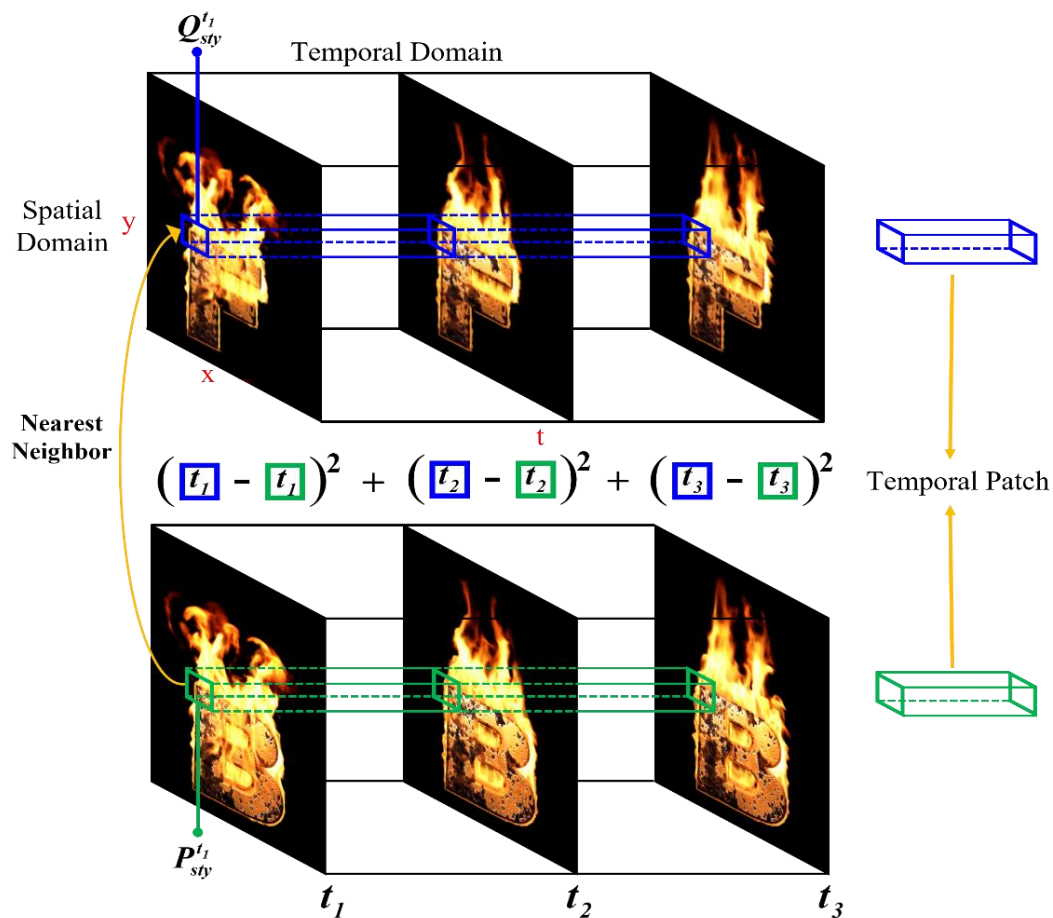
- **时序**问题：提升了时序稳定性，但仍然存在纹理抖动（尤其复杂动静混合纹理）
- **外观**问题：未解决

研究进展(2/3)

• 实现方法

■ 时序问题

全新的**共享匹配关系**概念：帧与帧之间共享匹配关系



- 共享匹配关系保证时序稳定性
- 采用多关键帧保证空间纹理连续性

时空相干项

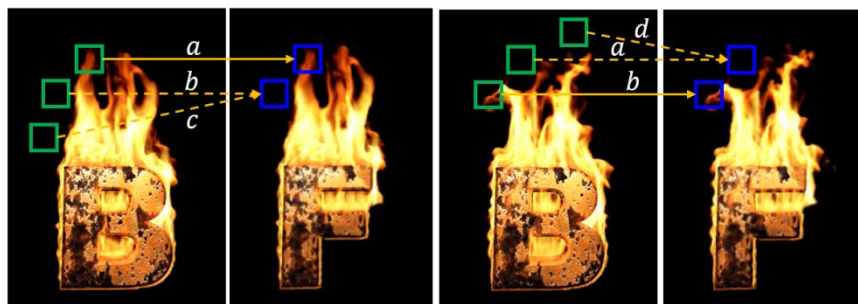
$$E_{st} = \sum_{t \in kf} D(P_{sty}^t - Q_{sty}^t)$$

研究进展(2/3)

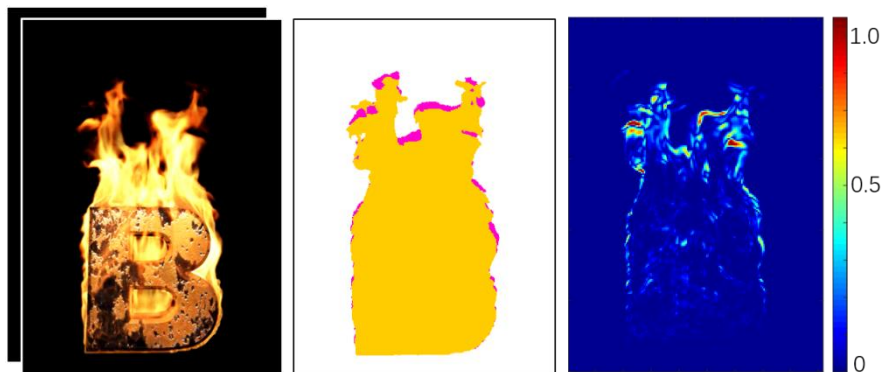
• 实现方法

关键帧提取

关键帧：相对上一帧发生明显强烈的粒子运动，新生产亡粒子数目高



(a) Source-target pairs at different frames



(b) $S_{sty}^{t-1,t}$

(c) $M_{sty}^{t-1,t}$

(d) \tilde{m}_t

粒子变化程度图

$$m_t = |g(S_{sty}^t) - g(S_{sty}^{t-1})|, 2 \leq t \leq N$$

粒子运动强度 $v_t = \sum_{p=1 \dots w \times h} \tilde{m}_t(p)$

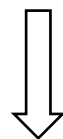
研究进展(2/3)

• 实现方法

目标函数 $E = \sum_{Q \in T} \min_{P \in S} (\lambda D(P_{text}, Q_{text}) + \blacksquare D(P_{sty}^t, Q_{sty}^t))$

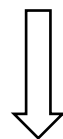
语义引导项

时空相干项



优化过程

共享匹配关系



应用于所有素材风格帧

目标风格动画

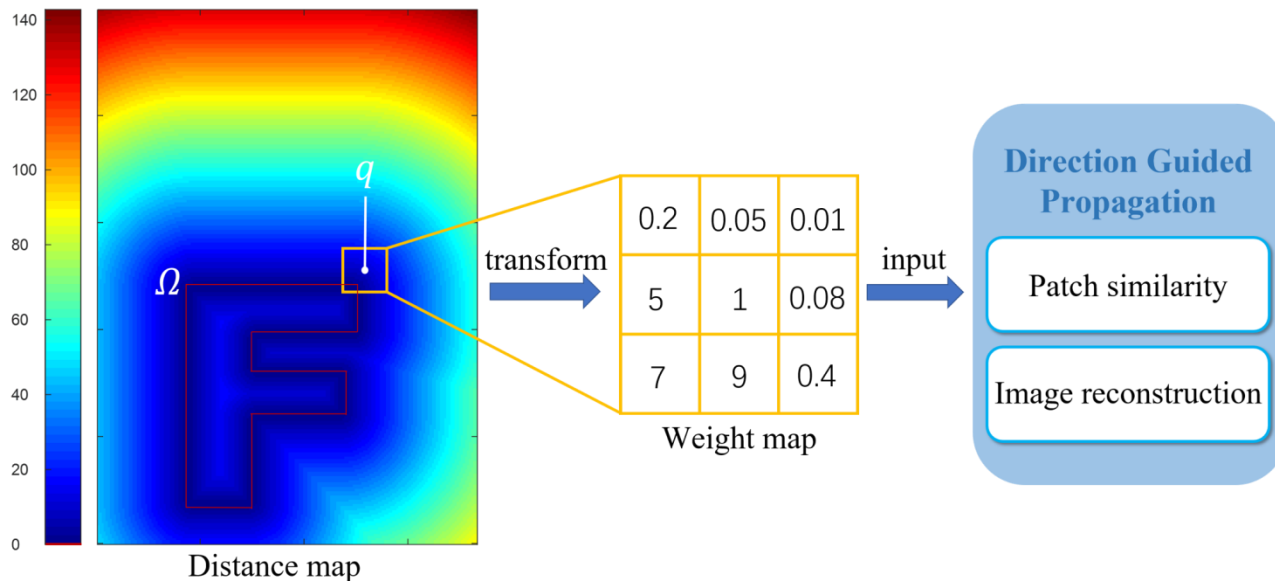
研究进展(2/3)

• 实现方法

■ 优化过程

有方向引导的深度匹配关系搜索算法 -> **外观问题**

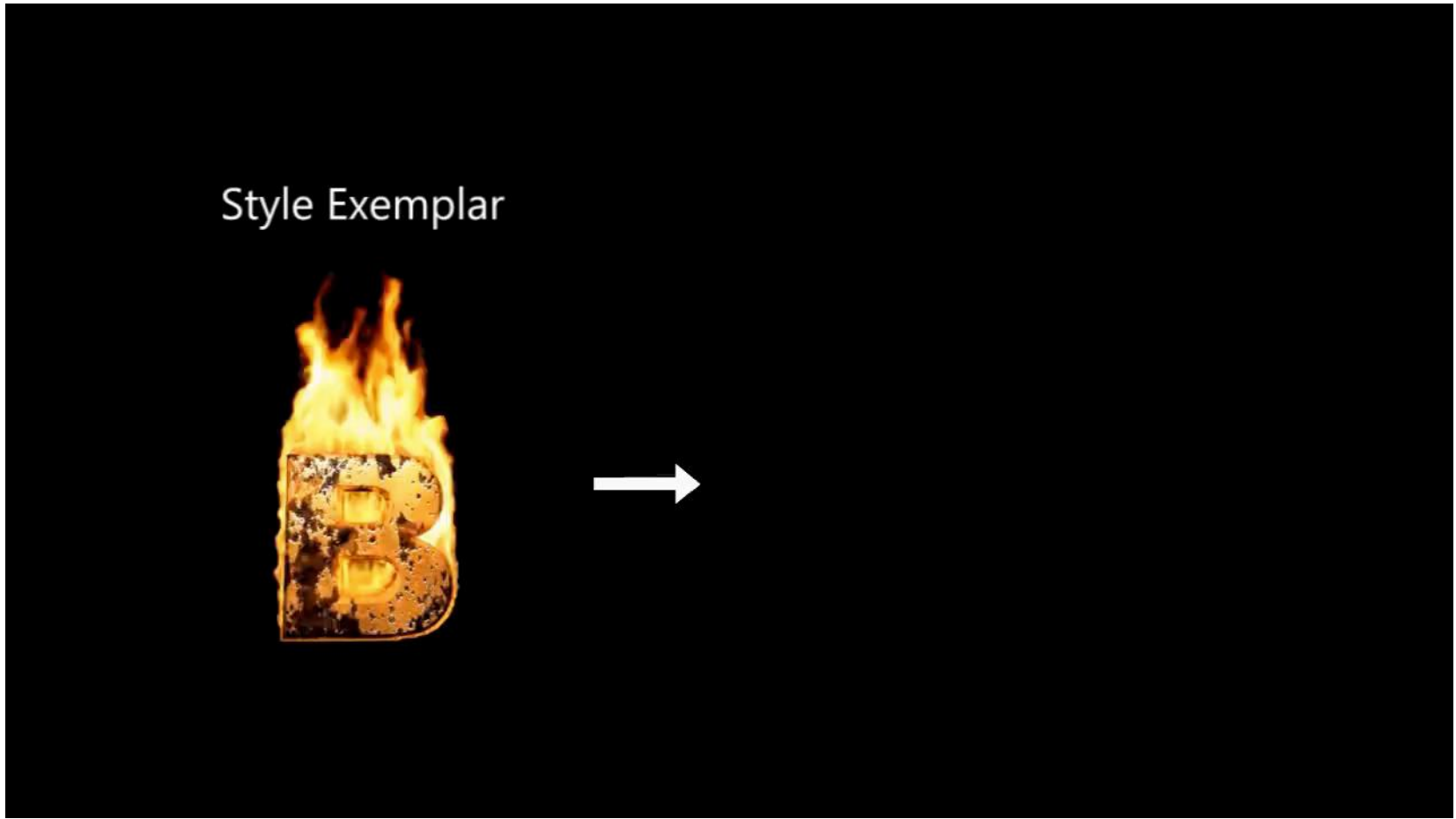
(1) 基于距离的权重地图 -> 有方向引导的纹理传播 (边缘到内部)



(2) 基于模拟退火的深度传播算法 -> 深度传播

研究进展(2/3)

- 实验结果



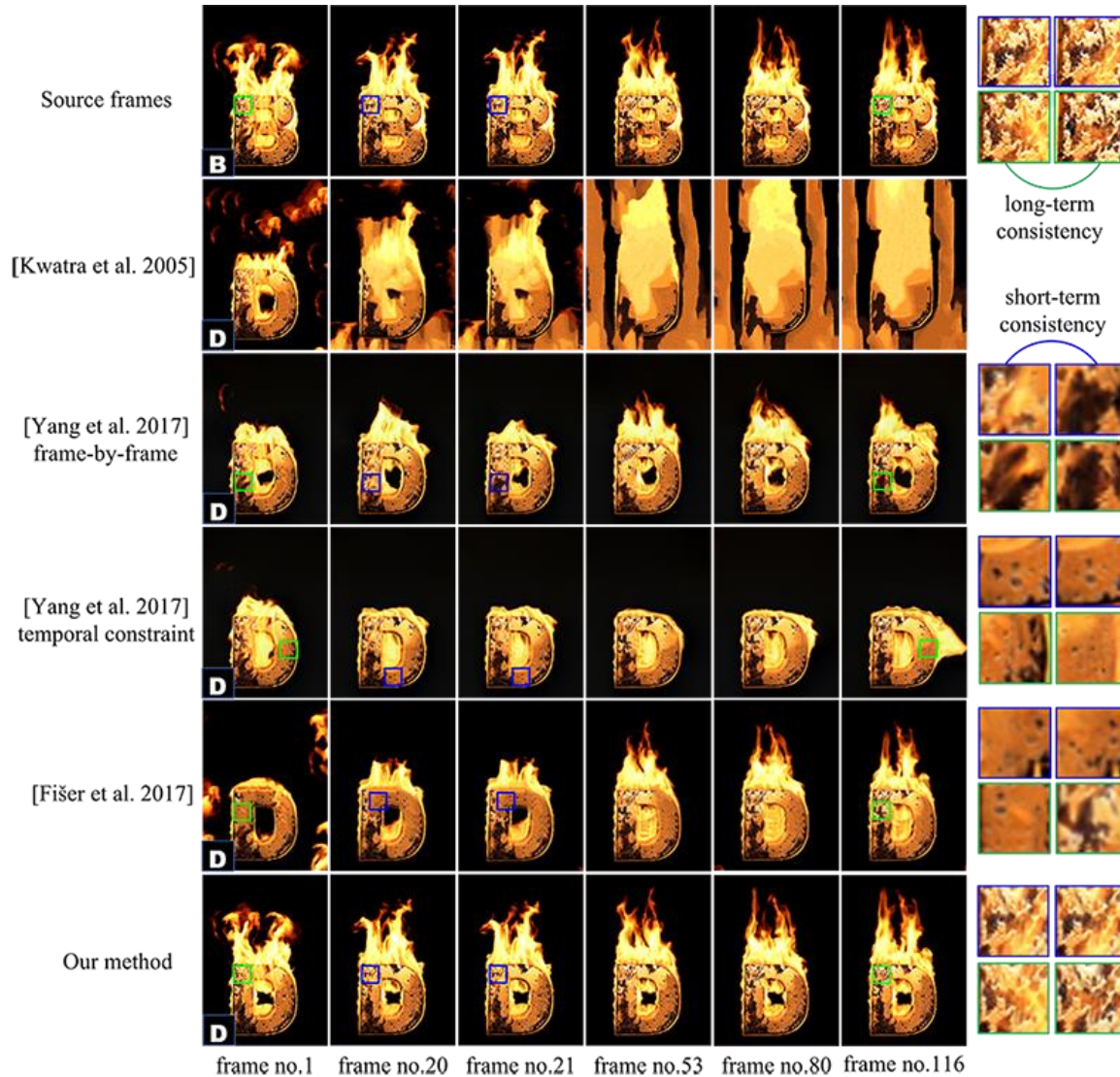
研究进展(2/3)

- 实验结果



研究进展(2/3)

• 实验结果 与其他方法对比



研究进展(2/3)

- **实验结果** 与其他方法对比

基于用户问卷调查的量化评估结果

指标	Flow ^[41]	T-Effects ^[67]	T-Effects ^[67] (w/ temp ^[34])	Animation Stylization ^[18]	本文方法
时序分数	17.5	7.5	45.0	57.5	97.5
外观偏好 (%)	0	5.75	4.75	6.25	83.25

研究进展(3/3)

Artistic Glyph Image Synthesis via One-Stage Few-Shot Learning

Y. Gao, Y. Guo, Z. Lian*, Y. Tang, J. Xiao

ACM Transactions on Graphics (TOG)
(presented at SIGGRAPH Asia 2019)



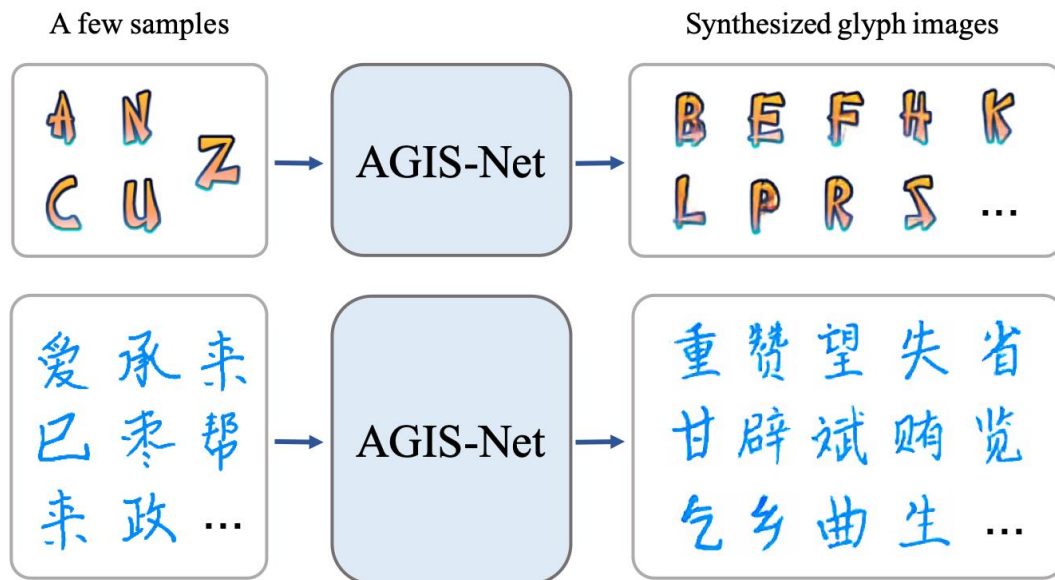
北京大學
PEKING UNIVERSITY

研究进展(3/3)

SIGGRAPH ASIA 2019

研究进展(3/3)

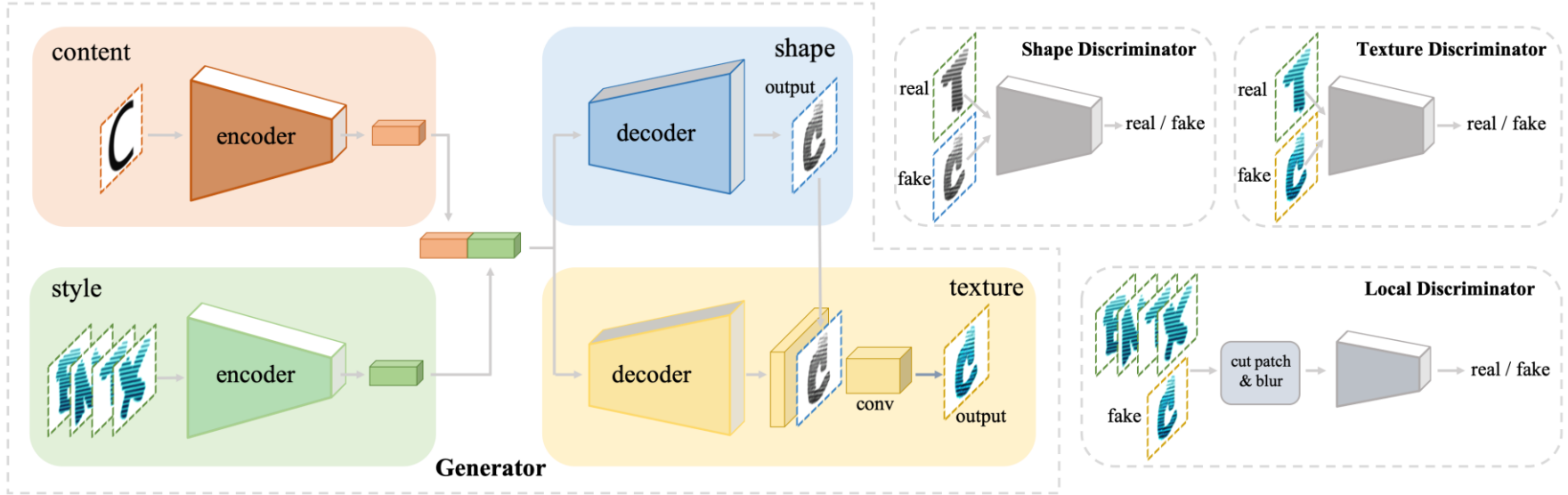
• 工作概览



设计包含特殊跳跃连接的形状和纹理风格解耦融合的双通路编解码器网络架构，并引入特殊的局部区域判别器来增强字形细节生成质量，首次实现基于小样本学习的任意大规模艺术风格字库的自动生成。

研究进展(3/3)

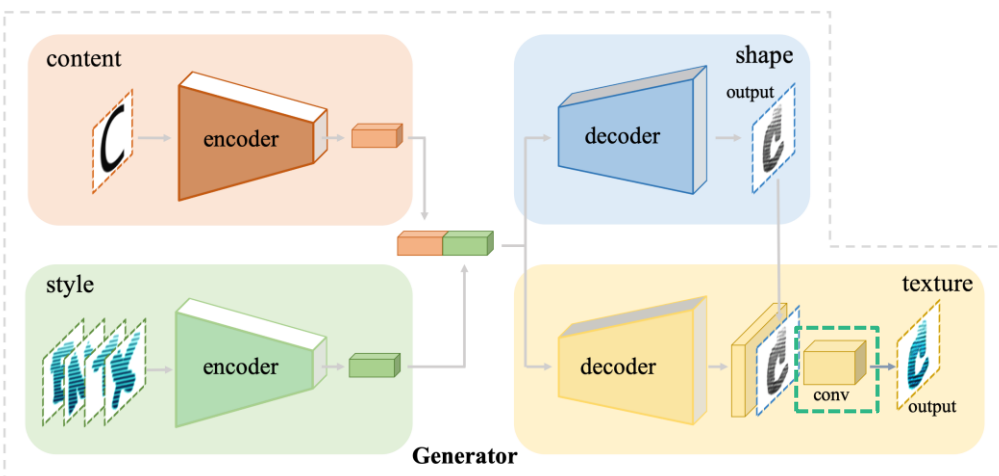
• 模型概览



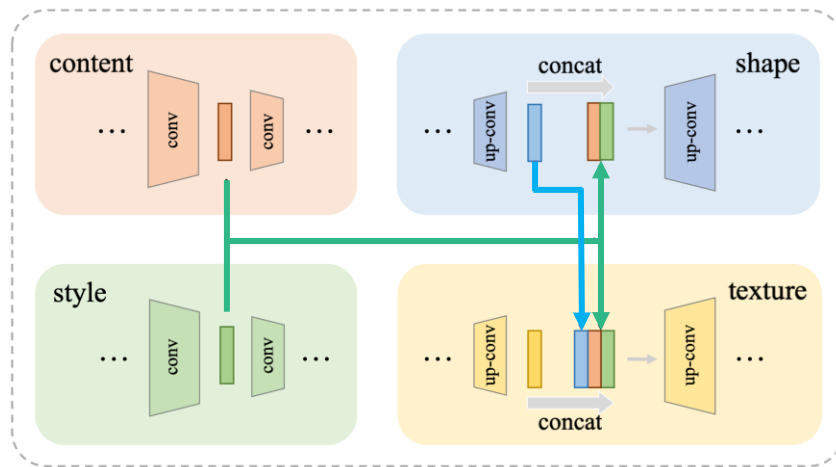
研究进展(3/3)

• 生成器结构

- Two collaboratively working encoder-decoder branches



- Special skip-connection

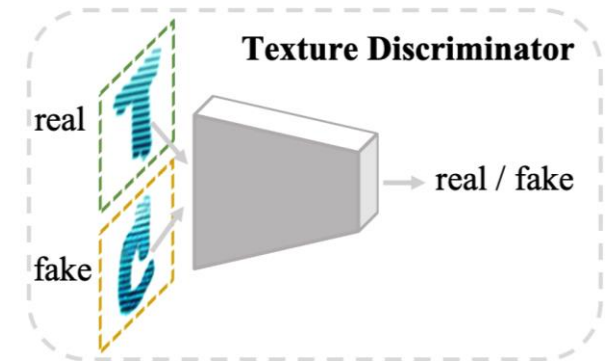
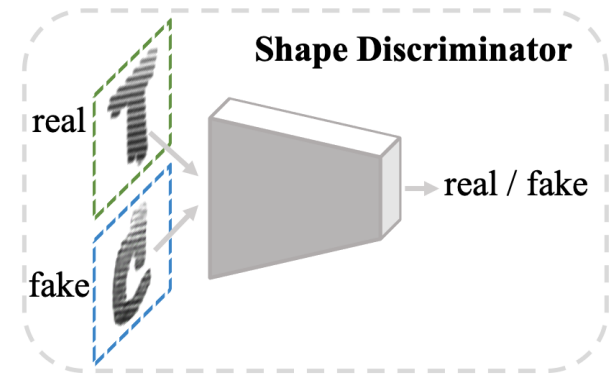


研究进展(3/3)

• 判别器结构

• Adversarial training:

- Shape Discriminator D_{sha}
- Texture Discriminator D_{tex}
- Local Discriminator

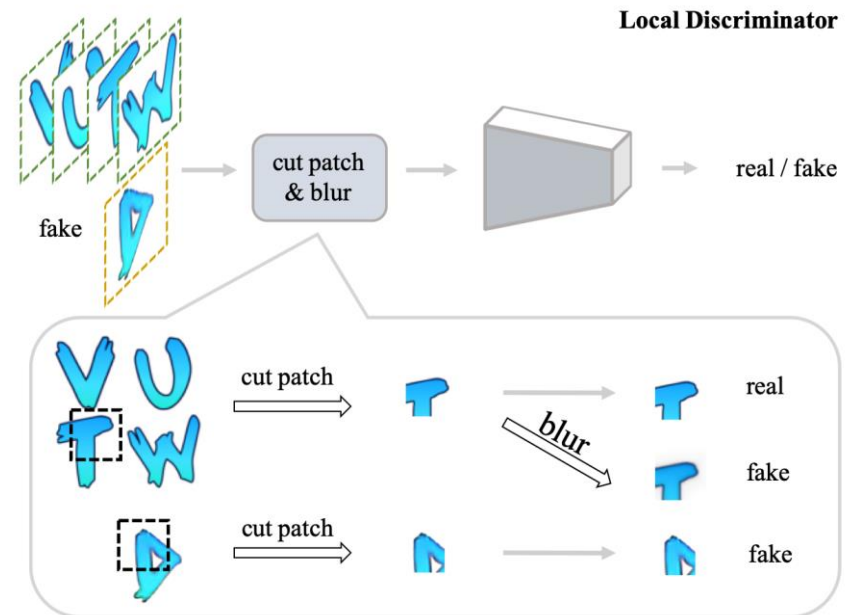


研究进展(3/3)

• 判别器结构

• Adversarial training:

- Shape Discriminator
- Texture Discriminator
- Local Discriminator D_{local}



研究进展(3/3)

• 目标函数

- Adversarial loss
- L_1 loss
- Contextual loss
- Local texture refinement loss

- Pre-train on synthetic dataset
- Fine-tune (few-shot learning) on specific artistic style

$$L(G) = L_{adv} + L_1 + L_{CX} + L_{local}$$

$$\min_G \max_{D_{sha}, D_{tex}, D_{local}} L(G, D_{sha}, D_{tex}, D_{local}).$$

研究进展(3/3)

• 实验结果

• Datasets

- Content glyph images



A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

For English fonts

赳 竦 忒 瑛 艾 民 我 凯 捷 患 躬 鸽 弟 承 白
缶 浆 戾 孑 哧 整 敖 臣 秦 举 辉 庚 顿 逮 辟
耒 聿 爰 彗 阉 奂 蛰 毕 朗 凰 鄱 花 丁 存 匙

For Chinese fonts

研究进展(3/3)

• 实验结果

• Datasets

- Synthetic artistic glyph images

English: 32,046 fonts with 26 characters (MC-GAN)

Chinese: 2,460 fonts with 639 characters



研究进展(3/3)

• 实验结果

• Competitors

- BicycleGAN [Zhu et al. 2017b]
- MS-Pix2Pix [Mao et al. 2019]
- MC-GAN [Azadi et al. 2018]
- TET-GAN [Yang et al. 2019]

• Evaluation metrics

- Inception Score (IS) [Salimans et al. 2016] ↑
- Fréchet Inception Distance (FID) [Heusel et al. 2017] ↓
- Structural similarity (SSIM) index ↑
- Pixel-level accuracy (pix-acc) ↑

研究进展(3/3)

• 实验结果

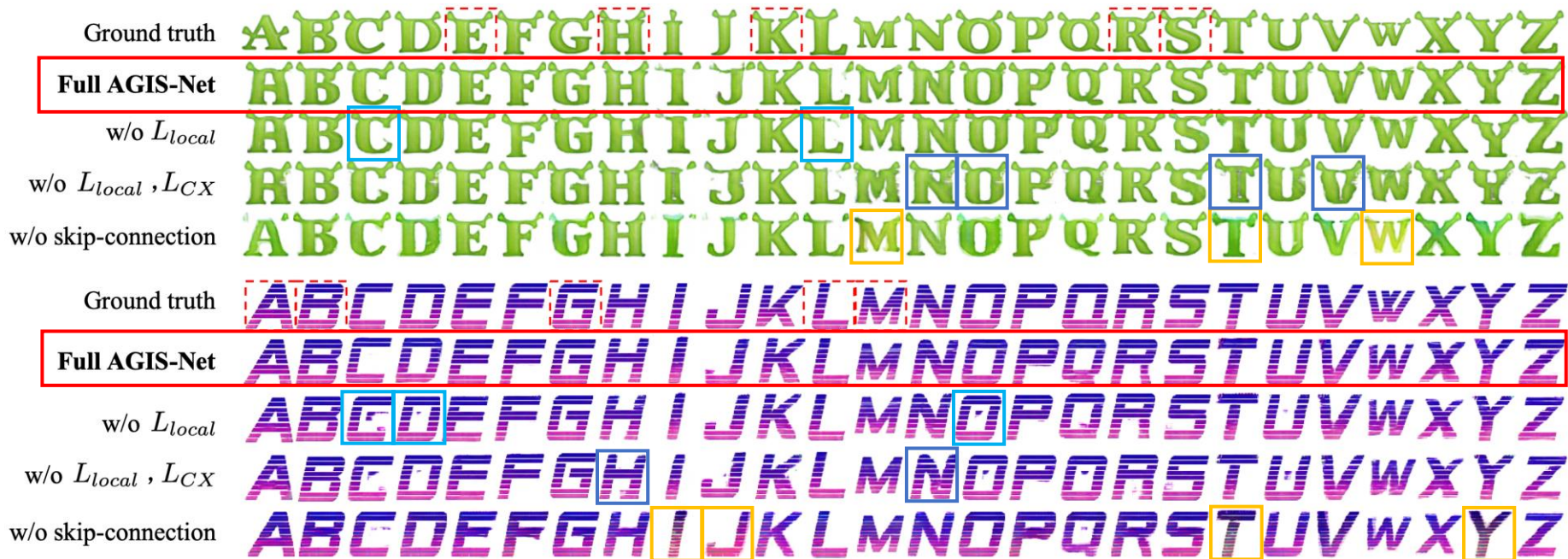
Pretraining Results

Ground truth													
BicycleGAN													
Ground truth													
MS-Pix2Pix													
Ground truth													
AGIS-Net													

研究进展(3/3)

• 实验结果

Ablation Study



SETTING	IS	FID	SSIM	PIX-ACC
w/o skip-connection	2.3365	109.555	0.6584	0.5876
w/o L_{local} and L_{CX}	2.4002	92.457	0.7063	0.6139
w/o L_{local}	3.2653	75.536	0.7219	0.6250
Full AGIS-Net	3.4046	74.567	0.7389	0.6241

研究进展(3/3)

• 实验结果

Comparison

Few-shot size $n = 5$
Style input size $m = 4$



研究进展(3/3)

- 实验结果

Comparison

MODEL	IS	FID	SSIM	PIX-ACC	USER PREFER.
TET-GAN	3.6378	193.043	0.4772	0.4322	0.0523
MC-GAN	3.7867	98.922	0.6620	0.5713	0.2463
AGIS-Net	3.8151	73.893	0.7217	0.6249	0.7014

研究进展(3/3)

• 实验结果

Few-shot size $n = 30$
Style input size $m = 4$

Comparison



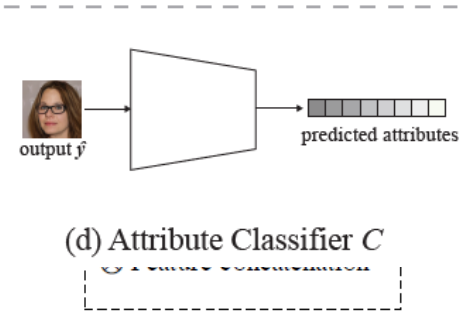
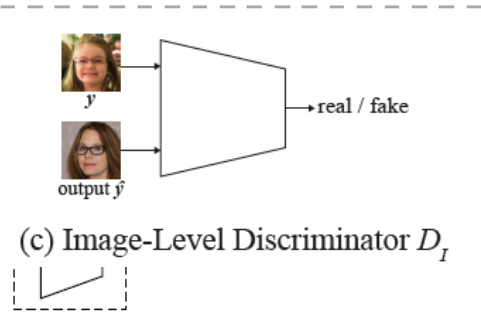
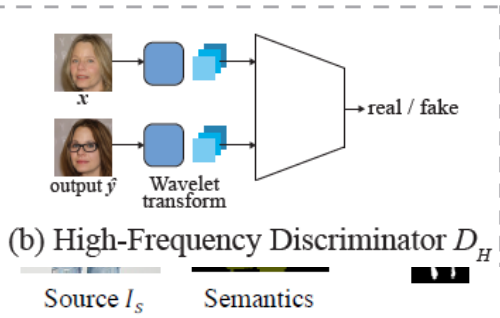
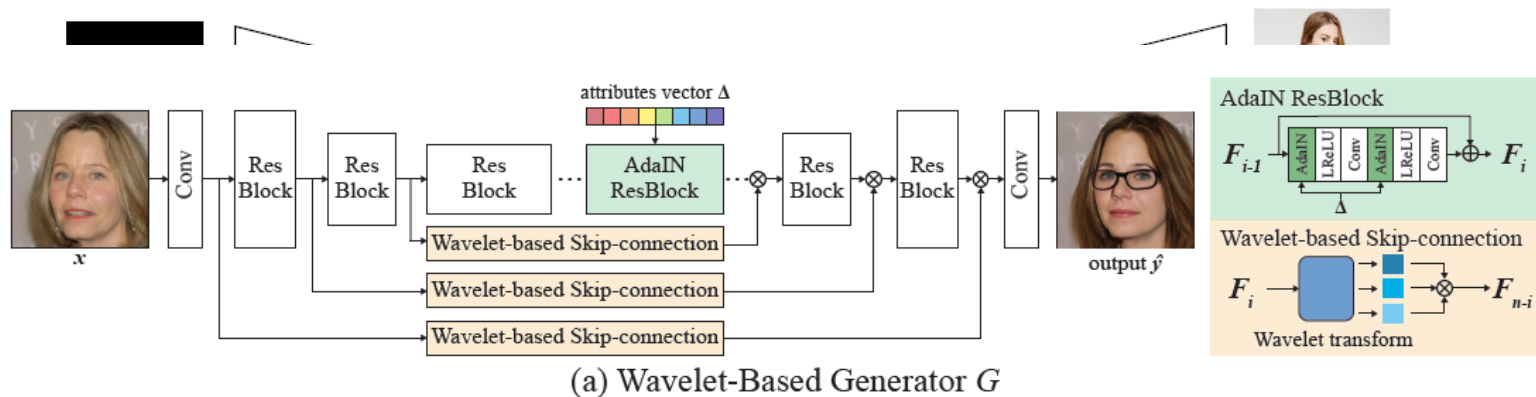
MODEL	IS	FID	SSIM	PIX-ACC
TET-GAN	2.3285	165.308	0.3937	0.4775
AGIS-Net	2.1122	70.875	0.6116	0.7035

研究进展(扩展应用)



扩展应用于人物图像、人脸图像、复杂场景图像合成

- ✓ 深度语义分割+自动属性分离 [CVPR 20](oral)
- ✓ 小波变换跳跃连接, 高频局部信息保留 [CVPR 21]



研究进展(扩展应用)

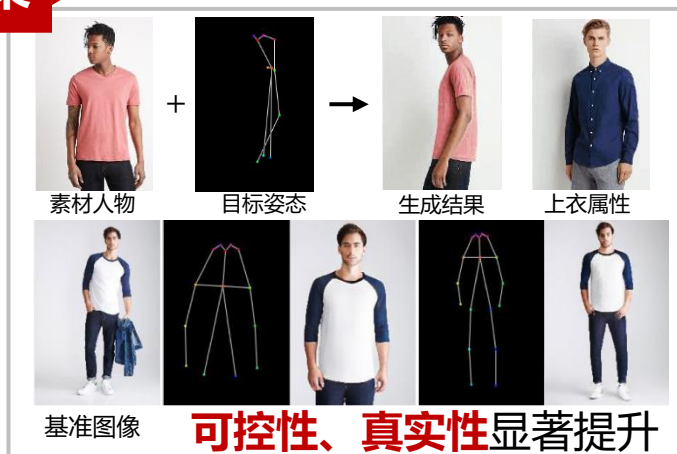


思路

扩展应用于人物图像、人脸图像、复杂场景图像合成

- ✓ 深度语义分割+自动属性分离 [CVPR 20](oral)
- ✓ 小波变换跳跃连接, 高频局部信息保留 [CVPR 21]

结果



原图

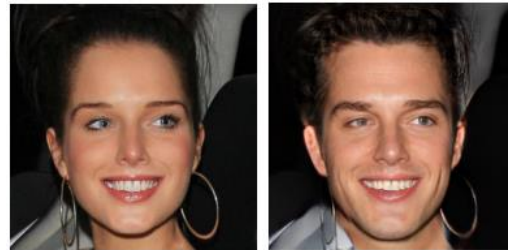
带眼镜

微笑



原图

合嘴



原图

性别改变

研究进展(论文列表)

- [1] Y. Gao, Y. Guan, **Z. Lian***, Y. Tang, J. Xiao. Artistic Glyph Image Synthesis via One-Stage Few-Shot Learning. ACM Transactions on Graphics (*TOG*). (presented at *SIGGRAPH Asia 2019*)
- [2] Y. Gao, F. Wei, J. Bao, S. Gu, D. Chen, F. Wen, **Z. Lian***. High-Fidelity and Arbitrary Face Editing, *CVPR 2021*
- [3] Y. Men, Y. Mao, Y. Jiang, W. Ma, **Z. Lian***. Controllable Person Image Synthesis with Attribute-Decomposed GAN. *CVPR 2020 (Oral)*
- [4] Y. Men, **Z. Lian***, Y. Tang, J. Xiao. DynTypo: Example-based Dynamic Text Effects Transfer. *CVPR 2019*
- [5] Y. Men, **Z. Lian***, Y. Tang, J. Xiao. A Common Framework for Interactive Texture Transfer. *CVPR 2018 (Spotlight)*
- [6] S. Yang, J. Liu, **Z. Lian**, Z. Guo. Awesome Typography: Statistics-Based Text Effects Transfer. *CVPR 2017*

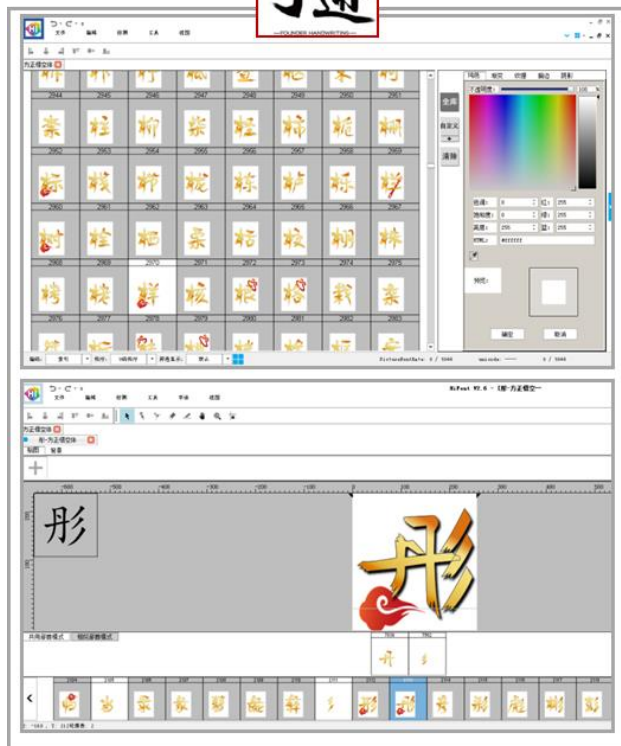
应用落地

方法
创新

局部感知合成

特效中文字库辅助设计系统

应用
落地



特效中文字库辅助设计系统

总结

1

文字是人类社会进步的核心要素，汉字是中华文明的**根基和载体**

2

字体相关技术是计算机图形学中的一个**重要研究内容**

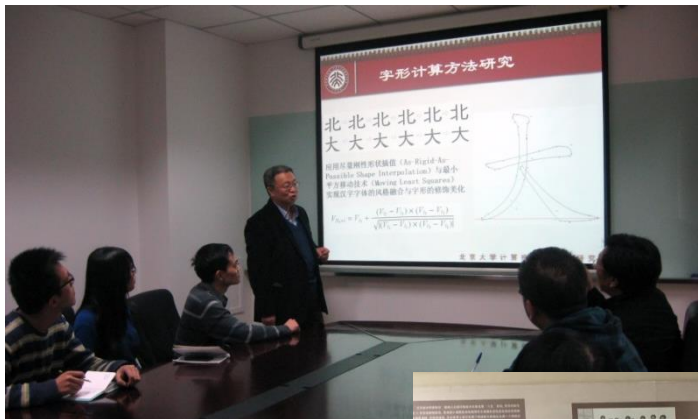
3

字体制作已从数字化**进入智能化时代**

4

深度模型**很强大**、传统方法**仍有效**

团队成员



☑ CCF-CV走进高校

汇报完毕 谢谢大家



北京大學
PEKING UNIVERSITY