主办 CCF计算机视觉专业委员会

COMPUTER VISION NEWSLETTER

CCF 计算机视觉专委会简报

01 2025

总第 43 期

CCF 计算机视觉专委会

COMPUTER VISION NEWSLETTER

计算机视觉专委会简报

2025 年第 01 期

总第 43 期

CCF 计算机视觉专委会

主　办　CCF 计算机视觉专业委员会

编委会

CONTENTS
简报目录

| 专 委 动 态

04 走进高校系列报告会
05 CCF CV 委员声音

| 科 技 前 沿

07 双曲核网络
13 三维人脸精细结构重建
21 基于三维高斯的动态人脸重建与几何纹理联合优化的语音驱动人脸

| 委 员 风 采

31 苏州科技大学胡伏原教授访谈
34 委员好消息

| 学 术 资 源

36 基于神经辐射场的渲染方法及开源代码
38 人体动作生成数据集好文推荐
40 好文推荐

| 海 外 学 者

43 征文通知

CCF 计算机视觉专委会
CCFCV.CCF.ORG.CN
CCFCVN@GMail.com

CCF-CV 走进高校系列报告会

第 143 期 北京工业大学

2025 年 1 月 3 日，由中国计算机学会主办，中国计算机学会计算机视觉专委会（CCF-CV）、北京工业大学联合承办的第 143 期 CCF-CV 走进高校系列报告会在北京工业大学理科楼 844 报告厅成功举行。本次报告会邀请了清华大学**高跃**长聘副教授、北京理工大学**袁野**教授、中科院自动化所**张兆翔**研究员、北京交通大学**景丽萍**教授等四位专家学者作特邀报告，北京工业大学师生齐聚一堂，积极参与这场学术盛宴，探讨计算机视觉的最新研究进展和未来发展趋势。出席本次报告会的领导有北京工业大学计算机学院院长**韩红桂**教授，专委会常务委员**毋立芳**教授，本期报告会执行主席是北京工业大学的**胡永利**教授和**王博岳**副教授，报告会由王博岳和简萌主持。

会议伊始，北京工业大学计算机学院院长韩红桂教授致辞，对出席论坛的各位专家学者表示热烈欢迎，简要介绍了学院在人工智能及相关领域的研究成果与发展目标，近年来在深度学习、模式识别、智能系统等方面取得了显著进展。最后，感谢专家学者分享最前沿的研究成果和宝贵的学术经验，以及一直以来对学院、研究院的支持与帮助。随后，CCF-CV 常务委员毋立芳教授代表专委致辞，首先感谢各位讲者、执行主席和领导对活动的大力支持。其次简单介绍了专委的各种活动、宣传平台等，欢迎大家承办、参与、关注专委的各项活动。最后预祝活动圆满成功！

会议首先由清华大学高跃老师、北京理工大学袁野老师、中科院自动化所张兆翔老师、北京交通大学景丽萍老师做主题报告，内容涵盖了超图计算、神经符号数据库、世界仿真器以及认知启发的高维数据概念学习等前沿研究方向。随后，四位专家与师生互动，共同探讨、交流，并对师生提出的问题做出详尽的回答，提出了许多有价值的学术见解，论坛现场气氛热烈。论坛最后，胡永利老师对报告会进行总结。

本期 CCF-CV 走进北京工业大学系列报告会围绕超图计算、神经符号数据库、世界仿真器、认知启发的高维数据概念学习等前沿技术展开，涵盖了计算机视觉及相关领域的多个维度。四位专家的报告不仅详尽阐述了最新的理论研究成果，还紧密结合了实际应用的案例，为与会师生呈现了一场精彩纷呈的学术盛宴。在报告过程中，专家们与师生积极互动，针对师生们提出的一系列问题，给予了详尽的解答，并分享了大量宝贵的学术观点与见解。现场氛围活跃，师生们踊跃提问，专家们耐心解答，充分体现了学术交流的深度与广度。此次报告会极大地鼓舞了师生们的科研动力，为师生搭建了一个难得的学习与交流平台，促进了知识与思想的碰撞与融合。

责任编委 黄岩

CCF-CV 委员声音

2025 年全国两会的春风，吹响了奋进的号角，大会汇聚了各方智慧与力量，为国家发展注入强劲动力。在这场关乎国计民生、民族未来的盛会中，中国计算机学会计算机视觉专委会（CCF-CV）的委员们带着对科技与教育的深刻洞察、对未来的热切期许，走进两会会场，发出专业强音，让我们一同聆听他们的智慧之声，感受科技赋能时代的磅礴力量。

谭铁牛院士强调，高水平研究型大学应充分发挥基础研究主力军、重大科技突破策源地和重大应用成果孵化器的重要作用，为高水平科技自立自强和中国式现代化建设提供有力支撑。

谭铁牛院士表示，增强基础研究原创能力，是提升科技支撑力的起始点。高水平研究型大学要发挥多学科的综合优势，聚焦国家重大需求，瞄准国际学科前沿，深耕基础研究，为技术创新提供理论基础和智力支持。

强化关键技术攻坚能力，是提升科技支撑力的着力点。高水平研究型大学要勇挑重担，在突破“卡脖子”关键核心技术上发挥国家战略科技力量的应有作用。

提高科技成果转化能力，是提升科技支撑力的落脚点。应持续深化校企校地合作，以科技创新和产业创新深度融合加速科研成果转移转化，助力培育新质生产力，更好服务国家和区域发展战略。

谭铁牛院士表示：“我们将不断强化基础研究、关键技术攻关和成果转移转化，持续增强学校的科技支撑力，服务高水平科技自立自强，为以中国式现代化全面推进强国建设、民族复兴伟业作出新的更大贡献。”

王亮委员在两会期间接受采访时表示，大模型技术的进步为通用人工智能（AGI）的实现带来了可能，但实现这一目标仍面临诸多挑战，要真正实现通用人工智能，仍需攻克诸多技术难题，例如在自动驾驶领域，仅依靠计算机视觉难以应对复杂环境，还需结合多模态解决方案。

王亮委员强调，推动人工智能的广泛应用需要高质量的多模态数据融合，并加快实现数据共享。他还指出，学术研究与产业应用的深度融合是推动人工智能加速落地的关键。需要在研发阶段加强产学研协同合作，缩短技术适配周期。

王亮委员的建议为人工智能的未来发展提供了重要思路：一方面，通过多模态数据融合和高质量数据标注提升大模型性能；另一方面，通过产学研深度融合加速技术落地，助力“人工智能+”战略的广泛实施。

高新波委员在两会期间接受记者采访时提出：“技术创新不一定非要沿着西方国家的轨迹亦步亦趋，后发者亦可定义技术框架。当前我国在 AI 大模型领域具有优势，应从技术创新、产业生态、应用场景等方面形成全方位、多层次的发展格局。”

高新波委员指出，我国 AI 大模型领域已具备政府支持、生态系统完善、市场需求广阔等优势，DeepSeek 的成功更是打破了“美国引领、中国跟随”的格局。他强调，未来产业竞争将转向生态体系的完备性，企业需在开放协作与技术主权之间寻求动态平衡。

在推动 AI 产业高质量发展方面，高新波建议从技术创新、产业生态、应用场景、人才培养和政策引导等多方面入手，形成全方位、多层次的发展格局。

在高校人才培养方面，高新波委员提出，高校应重构学科体系，打破学科壁垒，强化科教融汇和实践导向，培养复合型创新人才。他还建议通过国际合作与交流，培养具有国际视野的高技术人才。

责任编委 毋立芳

专题综述

双曲核网络

东南大学 方鹏飞

本文是东南大学团队针对双曲核网络方面的一系列研究成果，发表于ICCV 2021[1]、IJCV2023[2]。该系列工作研究的问题是围绕双曲空间（hyperbolic space），一种具有负曲率的曲面空间，定义有效的核函数并验证其有效性。近期研究表明，将数据嵌入至双曲空间作为其表征空间能够有效提升诸如小样本学习、零样本学习、度量学习等多种机器学习应用的性能。然而，由于双曲空间的曲面性质导致在该空间中的操作变得困难，例如，求解一系列双曲空间中的点的Fréchet均值需要迭代的算法。在欧式空间中，核方法不仅具备丰富的理论性质，同时具有强大的表征能力。本文研究双曲空间中的核方法，包含双曲正定核以及双曲不定核。这为表征学习带来两个好处：1-核化将使得表征兼具核机的强大表示能力以及双曲空间的编码能力；2-核空间（希尔伯特空间或克莱空间）丰富的结构可以简化多种双曲空间内的操作。具体而言，系列工作以嵌入方程思路展开，设计了多个双曲嵌入方程，并定义一系列双曲核函数。通过在一系列学习任务中验证提出的双曲核函数的有效性。

一、研究背景

在机器学习社区，欧式空间一直作为特征空间，用以编码诸如图像或文本等数据。其主要原因是高维欧式空间是我们所熟悉三维空间的自然概括，且在欧式空间中计算距离与相似度等算子较为简易。然而，将数据编码至欧式空间可能会损害或者扭曲数据的结构信息，从而丢失数据中固有的复杂空间几何信息。例如，欧式空间难以编码图结构数据中的层次信息[3]。近期，系列研究表明，相较于将数据嵌入至欧式空间，双曲空间作为表征空间能够刻画数据的层次结构，同时，该结构信息益于大量机器学习应用，包括文本蕴含[4]、机器翻译[5]、语言视觉推理[5]、图像分类和检索[6]、图数据分类[7]等任务。

双曲空间是一种具有恒定负曲率的曲面空间。该负曲率特性赋予双曲空间具有编码数据层次结构的性质。这是由于在双曲空间中，其体积会随着半径呈指数级增长[8]，从而提升了数据在该空间中的表示能力。尽管有一系列工作已经成功利用双曲空间进行推理[4-7]，但在非线性空间中进行数学运算等困难阻碍了其更广泛的使用。例如，在欧式空间中计算一系列数据的均值很简单，但双曲空间中的平均需要用Fréchet均值来近似。然而，计算Fréchet均值需要迭代算法，这使得算法或者神经网络的计算成本变得更加高昂[9,10]。这促使我们研究双曲空间中的核方法，以便能够无缝地利用核机以分析双曲数据。

核方法可以认为是一种衡量数据间相似度的度量。在欧式空间中，许多常见的核函数被定义为欧式距离（亦称为欧式空间的测地线距离）的函数。以常用的径向基核(RBF kernel)，记为$k(x,y)=\exp\left(-\xi d^2(x,y)\right)$，为例。有一种猜想是一旦知道测地线距离，就可以在曲面空间（双曲空间就是其中之一）中构造有效的正定核。不幸的是，情况并非如此，正如Jayasumana等人[11]和Feragen等人[12]所示，由于曲面空间与平坦的欧式空间不存在等距关系，故利用测地线距离定义的RBF核是非正定的。有趣的是，在曲面空间上定义正定核的难度现在被认为是机器学习领域中的一个悬而未决的问题[13]。

在本文中，我们使用庞加莱模型解决了在双曲空间中定义正定核的难题。在这里，我们提出了几个有效的正定双曲核，包括功能强大的通用核（universal kernel）。为此，我们首先利用引理构造一个有效的庞

加莱线性核。利用该引理，我们进一步为双曲几何定义了有效的庞加莱RBF核和庞加莱拉普拉斯核。同时，我们也提出了庞加莱二项式核。同时，针对上述核函数在神经网络优化中因调参等原因产生的繁琐问题，本文紧接着定义了庞加莱径向核，通过在庞加莱球中设计了一个通用公式，即对多核核函数进行加权和，获得双曲多核学习方法。本文在大量实验中验证上述双曲核函数的表达能力。

二、双曲核网络

2.1 前言：双曲空间

一个n维的双曲空间$\mathbb{H}^n$是一种具有恒定负曲率的黎曼流形[14]。庞加莱球是n维的双曲空间的一种建模形状，在庞加莱球中，所有的点（样本）均嵌入至一个n维的球体内。正式情况下，曲率为c的庞加莱模型定义为$\mathbb{D}_c^n = \{z \in \mathbb{R}^n : c\|z\| < 1\}$。它的黎曼度量定义为$g_c^{\mathbb{D}}(z) = \lambda_c^2(z) \cdot g^E$，其中$\lambda_c(z)$是适形因子，定义为$\lambda_c(z) = \frac{2}{1-c\|z\|^2}$，以及$g^E = I_n$是欧式度量。此外，我们借助莫比乌斯陀螺矢量空间（Möbius gyrovector space）便于矢量运算。具体而言，在庞加莱求中的莫比乌斯加法可定义为：

$$z_i \oplus_c z_j = \frac{(1 + 2c\langle z_i, z_j\rangle + c\|z_j\|^2)z_j + (1 - c\|z_i\|^2)z_j}{1 + 2c\langle z_i, z_j\rangle + c^2\|z_i\|^2\|z_j\|^2}$$

其中，$z_i, z_j \in \mathbb{D}_c^n$表示庞加莱球中的两个点。针对两个样本的测地线距离定义为：

$$d_c(z_i, z_j) = \frac{2}{\sqrt{c}} tanh^{-1}(\sqrt{c}\|-z_i \oplus_c z_j\|).$$

在黎曼集合中，一个点的切空间是一个内积空间，其包含了所有方向上与该点相切的向量。对于庞加莱球中的一个点，$z \in \mathbb{D}_c^n$，z点的切空间记为$T_z\mathbb{D}_c^n$。其中，指数映射定义为：

$$\mathcal{T}_z(p) = z \oplus_c (\tanh(\sqrt{c}\frac{\lambda_c(z)\|p\|}{2})\frac{p}{\|p\|})$$

其中$p \in T_z\mathbb{D}_c^n$。指数映射将在z的切平面中的一个向量p映射至庞加莱球$\mathbb{D}_c^n$。其逆过程称为对数映射，将庞加莱球中的一个向量$q \in \mathbb{D}_c^n$映射至z点的切空间，记为：

$$\mathcal{I}_z(p) = \frac{2}{\sqrt{c}}\lambda_c(z) tanh^{-1}(\sqrt{c}\|-z \oplus_c p\|)\frac{-z \oplus_c p}{\|-z \oplus_c p\|}$$

值得注意的是，指数映射与对数映射互为逆函数，满足：$\mathcal{T}_z(\mathcal{I}_z(p)) = p \in T_z\mathbb{D}_c^n$。在本文中，我们巧妙利用对数映射定义嵌入方程，以构建相应的双曲核函数。

2.2 双曲正定核

在本文中，我们首先利用双曲几何的切空间来定义一组有效的正定核。具体而言，我们首先提出一个映射，$f_{\mathbb{D}}(z): \mathbb{D}_c^n \to \mathbb{R}^n$，即将庞加莱球中一个向量映射至欧式空间，定义为：

$$f_{\mathbb{D}}(z) := tanh^{-1}(\sqrt{c}\|z\|)\frac{z}{\sqrt{c}\|z\|}$$

上式的物理意义是将庞加莱球中的向量z映射至原点的切空间。在原点的切空间，每一个向量$f_{\mathbb{D}}(z)$不仅能估计其在原庞加莱球中的原始向量z，同时，我们ICCV21[1]的论文中通过证明曲线长度等效定理揭示在两个空间中优化度量的等效性。基于该映射，本文提出了一系列正定核，具体如下：

庞加莱正切核：在欧式空间中，最简单的正定核为线性核。我们在双曲空间中也定义相应的线性核，记为：$k^{\tan}(z_i, z_j) = \langle f_{\mathbb{D}}(z_i), f_{\mathbb{D}}(z_j)\rangle$。该核可以理解在庞加莱球切空间中定义的线性核，故将其命名为庞加莱正切核。该核没有参数，适合对原型的快速验证。

庞加莱径向基核：本文通过利用定义的映射方程$f_{\mathbb{D}}(z)$以及在该空间的度量$d_e(\cdot,\cdot)$，提出庞加莱径向基核，表示为：$k^{\mathrm{rbf}}(z_i, z_j) = \exp\left(-\xi\|f_{\mathbb{D}}(z_i) - f_{\mathbb{D}}(z_j)\|^2\right)$。在ICCV21[1]文中，通过证明$\|f_{\mathbb{D}}(z_i) - f_{\mathbb{D}}(z_j)\|^2$是负定性质进而证明庞加莱径向基核$k^{\mathrm{rbf}}(z_i, z_j)$为正定核。庞加莱径向基核具有强大的通用估计（universal approximation）能力。

庞加莱拉普拉斯核：另一个常用且具备通用估计能力的核称之为拉普拉斯核。在双曲空间中，本文定义庞加莱拉普拉斯核，记为：$k^{\mathrm{lap}}(z_i, z_j) = \exp(-\xi\|f_{\mathbb{D}}(z_i) - f_{\mathbb{D}}(z_j)\|)$。其正定性的证明过程与庞加莱径向核证明过程类似。

庞加莱二项式核：除了上述指数型核函数，本文继续构造庞加莱二项式核函数，定义为$k^{\mathrm{bin}}(z_i, z_j) = (1 - \langle f_{\mathbb{D}}(z_i), f_{\mathbb{D}}(z_j)\rangle)^{-\alpha}$。该核函数通过泰勒分解获得非负性

完全泰勒级数证明其正定性，并且具备通用估计能力。

庞加莱径向核：上述核函数为单一的核函数，其在优化的过程中常常受数据、模型等影响，需要调试其参数。该问题启发我们在双曲空间中研究多核学习（multiple-kernel learning，MKL）方法，使得核函数能够根据数据或者模型本身学习相应的核函数。具体而言，基于$f_{\mathbb{D}}(z)$，我们提出一个新的映射函数$g_{\mathbb{D}}(z):=\frac{f_{\mathbb{D}}(z)}{\|f_{\mathbb{D}}(z)\|}$，并提出庞加莱径向核：$k^{\mathrm{rad}}(z_i,z_j)=\sum_{m=-2}^{\infty}a_m k_m^{\cos}(z_i,z_j)$，其中$k_m^{\cos}(z_i,z_j)$表示为余弦核。

三、实验结果

3.1 小样本学习

小样本学习（Few-shot Learning）是一种新的学习范式，其通过少量样本学习数据的表征空间，并使得该表征空间能够有效适用于未见过的测试数据[15-17]。通常而言，小样本学习通过元学习的方式训练神经网络。具体而言，每次迭代均会采样一个情节的数据用以训练网络，该训练方式称之为 N-way K-shot，并通过该训练方式实现在每个情节中对 N 个类别的识别任务。小样本学习主要基于度量学习思想展开，用以学习同时适用于可见样本与未见样本的度量空间。在匹配网络(Matching Network) [15]中，模型通过学习与样本相关度量以确定查询样本的类别。受匹配网络启发，原型网络(Prototypical Network, ProtoNet)[16]设计了与类别相关的度量方法，即每个类别的所有样本均被认为是对该类别的描述。在关系网络 (Relation Network) [17]中，

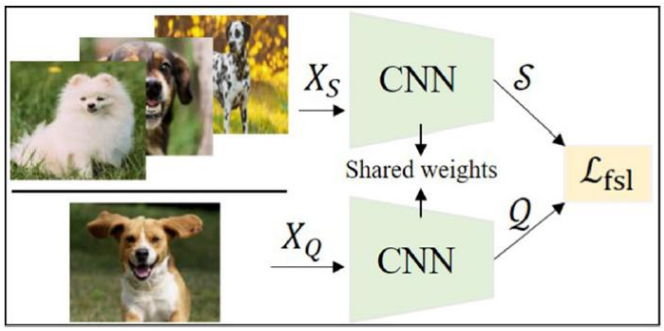


图 1　原型网络示意图

通过研究显式地建模支持样本与查询样本间的非线性相似度关系，使得所学习的隐度量空间能够自适数据本身的分布情况。

在我们的实验中，采用 5-way 1-shot 或者 5-way 5-shot 以评估模型。该实验采用原型网络模型（ProtoNet）[15]用以训练神经网络特征提取器。图 1 为原型网络的示意图。为验证所提出的双曲核函数的优越性，我们采用具有双曲表征能力原型网络为基线，记为 Hyper ProtoNet[6]。同时，我们采用 Conv-4 以及 ResNet-18 作为模型的骨干网络，用以提取数据的特征。同时，我们采用 tiered-ImageNet 以及 FC100 数据集用以验证模型提出的双曲核函数的效果。实验结果如表 1 所示。

如表 1 所示，本文所提出的双曲核函数较基线模型具有显著提升。在 tiered-ImageNet 数据集上，最简单的庞加莱正切核在 Conv-4 和 ResNet-18 骨干网络分别提升 (0.29%&2.46%) / (1.03%&1.56%)。在 FC100

Model	Backbone	*tiered*-ImageNet		FC100	
		5-way 1-shot	5-way 5-shot	5-way 1-shot	5-way 5-shot
Hyper ProtoNet† (Khrulkov et al., 2020)	Conv-4	54.44 ± 0.23	71.96 ± 0.20	37.59 ± 0.19	51.76 ± 0.19
Poincaré tangent kernel	Conv-4	54.73 ± 0.22	74.37 ± 0.18	37.66 ± 0.17	52.29 ± 0.18
Poincaré RBF kernel	Conv-4	57.78 ± 0.23	76.11 ± 0.18	38.93 ± 0.18	54.40 ± 0.18
Poincaré Laplace kernel	Conv-4	57.33 ± 0.22	76.48 ± 0.18	37.99 ± 0.17	53.54 ± 0.18
Poincaré binomial kernel	Conv-4	56.72 ± 0.22	75.87 ± 0.18	38.32 ± 0.18	53.50 ± 0.18
Poincaré radial kernel	Conv-4	**57.96 ± 0.22**	**76.87 ± 0.18**	**39.24 ± 0.17**	**54.82 ± 0.18**
Hyper ProtoNet† (Khrulkov et al., 2020)	ResNet-18	62.28 ± 0.23	74.50 ± 0.21	40.64 ± 0.20	52.50 ± 0.30
Poincaré tangent kernel	ResNet-18	63.31 ± 0.23	76.06 ± 0.23	42.18 ± 0.26	54.32 ± 0.32
Poincaré RBF kernel	ResNet-18	64.52 ± 0.22	76.82 ± 0.21	43.84 ± 0.23	56.01 ± 0.30
Poincaré Laplace kernel	ResNet-18	64.38 ± 0.22	77.16 ± 0.21	43.22 ± 0.23	55.47 ± 0.30
Poincaré binomial kernel	ResNet-18	64.12 ± 0.23	76.44 ± 0.23	42.60 ± 0.24	55.08 ± 0.32
Poincaré radial kernel	ResNet-18	**65.33 ± 0.21**	**77.48 ± 0.20**	**44.12 ± 0.20**	**56.28 ± 0.26**

† indicates the network was self-implemented. 1st / 2nd best in "bold" / "(underline)"

表 1　双曲核函数在小样本学习中的实验结果

数据集上，最简单的庞加莱正切核在 Conv-4 和 ResNet-18 骨干网络分别提升（0.07%&0.53%）/（1.54%&1.82%）。其余的双曲单核函数均有进一步的提升。其中，庞加莱径向基核在 Conv-4 为骨干网络时表现整体最优，而庞加莱拉普拉斯核在 ResNet-18 为骨干网络时表现整体最优。相较于双曲单核函数，多核学习方法也在小样本学习中展示其优越性，庞加莱径向核在 tiered-ImageNet 以及 FC100 中均获得最优性能。其在 tiered-ImageNet 数据集对 Conv-4 和 ResNet-18 骨干网络分别提升（3.52%&4.91%）/（3.05%&2.98%）；在 FC100 数据集对 Conv-4 和 ResNet-18 骨干网络分别提升（1.65%&3.06%）/（3.48%&3.78%），进一步揭示多核学习方法对提升双曲核函数表征能力的优越性。

双曲核函数的优越性在零样本学习、度量学习、行人重识别等机器学习应用中均有验证，感兴趣的读者请参考文献[1,2]。

四、总结

本文在双曲空间提出了一系列正定核，用以将双曲空间中的表征映射至希尔伯特空间。为了定义此类核函数，我们利用庞加莱球的恒等切空间（即庞加莱球原点的切空间），并进一步在恒等切线空间中定义有效的正定核函数。所提出的核函数包括功能强大的通用核函数（即庞加莱径向基核函数、庞加莱拉普拉斯核函数、庞加莱二项式核函数和庞加莱径向核函数）。我们在小样本学习等任务中评估了所提出双曲核函数的有效性。实验结果表明，这些核函数对提升双曲空间中的表征取得积极效果。未来的工作包括探索双曲负定核函数在相关机器学习应用中的有效性。

五、前期相关工作

核方法作为机器学习领域的重要理论框架，其系统性研究已在支持向量机（SVM）、主成分分析（PCA）及聚类算法等[1]经典模型中展现出显著优势。其理论的核心在于通过将原始数据映射至高维（甚至无限维）的再生核希尔伯特空间（Reproducing Kernel Hilbert Space, RKHS），在该空间中用线性模型解决原线性不可分问题。为规避显式求解高维映射的计算复杂度，核技巧利用核函数直接计算再生核希尔伯特空间内的样本相似度。基于此理论，研究者相继提出了多项式核、径向基核及拉普拉斯核等经典核函数形式[18]。为改善单核方法的适用性，多核学习（Multiple Kernel Learning, MKL）[19,20]框架应运而生。在多核学习中，核函数为基核的凸组合，权重从数据中习得，使得所学习到的核机器能够最大限度匹配数据的内在结构[19,21]。

近年来，为增强结构化数据表征能力，将核方法扩展至非线性几何空间的研究方向备受关注。在非欧几里得几何空间定义有效正定核（Positive Definite Kernel, PD Kernel）的通用策略是采用适配的距离度量。Jayasumana 等人[22]开创性地建立了对称正定矩阵空间的高斯核理论体系，随后该成果被进一步拓展至 Grassmann 流形[11]。Harandi 等人[23]系统探索了 Grassmann 流形上的核函数构造方法，提出基于等效嵌入映射的正定 Grassmannian 核。而 Le 与 Yamada[24] 将使用 Fisher 信息度量的核函数用于持续图。Jayasumana 等人[25]则发展了一系列紧致流形（包括n维球面、Grassmann 流形及形状流形）的径向基核统一框架。在现有研究中，Cho 等人[26]提出的双曲空间中的支持向量机方法与本文工作最具相关性，其虽尝试在双曲几何中引入核方法以构建非线性决策边界，但所提出的核函数因缺乏正定性导致两个本质缺陷：其一，不定核（Indefinite Kernel）不是一致逼近核，违背了通用逼近性质[27]；其二，不定核因其需要稳定损失值，导致训练困难[28]。

本文针对上述理论的局限，构建了双曲几何空间的正定核理论框架。作为对不定核的理论补充，本文通过双曲空间的核化过程，将双曲数据嵌入高维（可能无限维）希尔伯特空间，使得所得表征能够充分受益于核机器的理论优势。本文后续章节将详细阐述理论推导过程，并在多组具有挑战性的实际应用中评估算法的有效性。

责任编委 魏秀参

参考文献

[1] P. Fang, M. Harandi, and L. Petersson. Kernel methods in hyperbolic spaces. In ICCV 2021.

[2] P. Fang, M. Harandi, Z. Lan and L. Poincaré kernels for hyperbolic representations. IJCV, 2023.

[3] Q. Liu, M. Nickel, and D. Kiela. Hyperbolic graph neural networks. In NeurIPS 2019.

[4] O. Ganea, G. Bécigneul, and T. Hofmann. Hyperbolic neural networks. In NeurIPS 2018.

[5] C. Gulcehre, M. Denil, M. Malinowski, A. Razavi, R. Pascanu, K. Hermann, P. Battaglia, V. Bapst, D. Raposo, A. Santoro, and N. Freitas. Hyperbolic attention networks. In ICLR 2019.

[6] V. Khrulkov, L. Mirvakhabova, E. Ustinova, I. Oseledets, and V. Lempitsky. Hyperbolic image embeddings. In CVPR 2020.

[7] Q. Liu, M. Nickel, and D. Kiela. Hyperbolic graph neural networks. In NeurIPS 2019.

[8] M. Hamann. On the tree-likeness of hyperbolic spaces. Mathematical Proceedings of the Cambridge Philosophical Society, 2017.

[9] H. Karcher. Riemannian center ofmass and mollifier smoothing. In Communications on Pure and Applied Mathematics, 1977.

[10] A. Lou, I. Katsman, Q, Jiang, S. Belongie, S. Lim, and C. Sa. Differentiating through the fréchet mean. In ICML 2020.

[11] S. Jayasumana, R. Hartley, M. Salzmann, H. Li, and M. Harandi. Kernel methods on Riemannian manifolds with gaussian RBF kernels. TPAMI, 2015.

[12] A. Feragen, F. Lauze, and S. Hauberg. Geodesic exponential kernels: When curvature and linearity conflict. In CVPR 2015.

[13] A. Feragen, and S. Hauberg. Open problem: Kernel methods on manifolds and metric spaces. What is the probability of a positive definite geodesic exponential kernel. In CoLT 2016.

[14] P. Absil, R. Mahony, and R. Sepulchre. Optimization algorithms on matrix manifolds. Princeton University Press. 2007.

[15] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, and H. Wieretra. Matching networks for one shot learning. In NeurIPS 2016.

[16] J. Snell, K. Swersky, and R. Zemel. Prototypical networks for few-shot learning. In NeurIPS 2017.

[17] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. Torr, and T. Hospedales. Learning to compare: Relation network for few-shot learning. In CVPR 2018.

[18] T. Hofmann, B. Schölkopf, and A. J. Smola. Kernel methods in machine learning. The Annals of Statistics, 2008.

[19] A. Rakotomamonjy, F. R. Bach, S. Canu, and Y. Grandvalet. SimpleMKL. JMLR, 2008.

[20] G. R. G. Lanckriet, N. Cristianini, P. Bartlett, L. E. Ghaoui, and M. I. Jordan. Learning the kernel matrix with semidefinite programming. JMLR, 2004.

[21] T. Wang, L. Zhang, and W. Hu. Bridging deep and multiple kernel learning: A review. Information Fusion, 2021.

[22] S. Jayasumana, R. Hartley, M. Salzmann, H. Li, and M. Harandi. Kernel methods on the riemannian manifold of symmetric positive definite matrices. In CVPR 2013.

[23] M. T. Harandi, M. Salzmann, S. Jayasumana, R. Hartley, and H. Li. Expanding the family of grassmannian kernels: An embedding perspective. In ECCV 2014.

[24] T. Le, and M. Yamada. Persistence fisher kernel: A Riemannian manifold kernel for persistence diagrams. In NeurIPS 2018.

[25] S. Jayasumana, R. Hartley, M. Salzmann, H. Li, and M. Harandi. Optimizing over radial kernels on compact manifolds. In CVPR 2014.

[26] H. Cho, B. DeMeo, J. Peng, and B. Berger. Large-margin classification in hyperbolic space. In ICML 2019.

[27] C. A. Micchelli, Y. Xu, and H. Zhang. Universal kernels. JMLR, 2006.

[28] C. S. Ong, X. Mary, S. Canu, and A. J. Smola. Learning with non-positive kernels. In ICML 2004.

方鹏飞

东南大学计算机科学与工程学院教授，博士生/硕士生导师。分布于 2017 年与 2022 年在澳大利亚国立大学获得硕士与博士学位，随后于澳大利亚蒙纳士大学担任博士后研究员。2023 年加入东南大学，任副教授，教授。长期从事机器学习、计算机视觉领域研究工作。

Email:fangpengfei@seu.edu.cn

热点追踪

三维人脸精细结构重建

王子都 朱翔昱 雷震
中国科学院自动化研究所

本文主要介绍中国科学院自动化研究所生物识别与安全技术研究中心在三维人脸精细结构重建领域的最新研究成果，包括发表在CVPR 2024的3DDFA-V3[1]和ACM MM 2024的S2TD-Face[2]，并围绕其技术细节和创新点展开讨论。

一、研究背景

三维人脸重建作为计算机视觉与图形学领域的一个重要的方向，在生物识别、影视娱乐、人机交互、医疗美容等具有广阔的应用前景。近年来，基于深度学习的三维人脸重建方法利用数据驱动和模型先验，能够从非受控条件下的人脸图片中建模三维人脸几何形状和纹理信息，典型方法如 3DDFA 系列[3][4]、DECA[5]、Deep3D[6]、HRN[7]等。然而，三维人脸的精细结构重建在许多应用中仍然面临巨大挑战，并具有不可替代的重要性。例如，在数字人驱动中，精细的人脸形状结构建模直接影响虚拟人脸的真实感和自然度；在精神状态分析中，微表情和面部细节的准确捕捉对心理状态的推测具有重要价值。为进一步推动这一领域的发展，本文探讨了三维人脸精细结构重建中的两个具体研究工作，旨在抛砖引玉，助力相关领域的实际应用。

(a)

(b)

图 1 (a) 现有方法难以重建闭眼、歪嘴等极端表情。(b) 三维误差有时难以精确描述二维区域的对齐情况。

二、部件分割引导的人脸细节捕捉

2.1 背景介绍

从二维图像中重建三维人脸是计算机视觉研究的一项关键任务，研究人员通常依赖三维可变形模型（3DMM）进行人脸重建，以定位面部特征和捕捉表情。如图 1 所示，现有的方法往往难以准确重建出如闭眼、歪嘴、皱眉等极端表情，并且三维误差有时难以精确描述二维区域的对齐情况。为了增强 3DMM 对极端表情的捕捉和重建能力，3DDFA-V3 从训练策略和数据策略两个角度进行研究，以人脸分割为研究切入点，使用人脸部件分割的几何信息作为监督信号，设计损失函数，显著加强了对形状的约束，同时，3DDFA-V3 设计了可靠的表情生成方法，能够大批量、可控地生成难以获取的极端表情人脸图像。

2.2 研究内容

研究内容 1：训练策略的研究

人脸部件分割能够以像素级的精度为每个面部特征提供准确的定位。相比常用的关键点信息，部件分割提供了覆盖整个图像的更密集的标签；相比纹理信息，部件分割不易受到光照等因素的干扰。如图 2 所示，3DDFA-V3 的总体思路是利用分割信息来直接指导三维形变，进一步增强对人脸形状的约束。

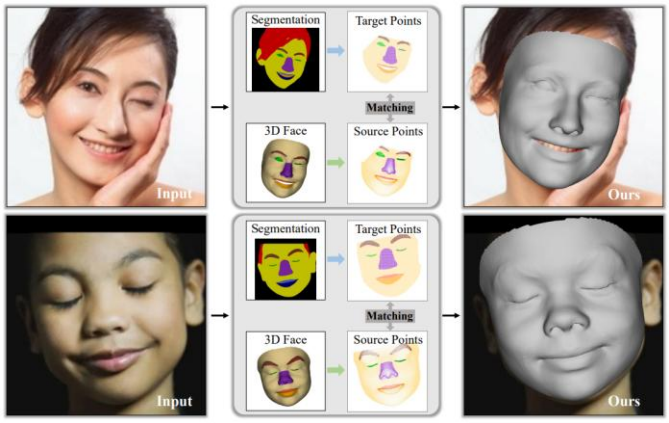


图 2　3DDFA-V3 的总体思路

现有利用分割信息来拟合三维人脸的尝试，依赖于可微渲染器生成预测的面部区域（如眼睛、眉毛、嘴唇等）轮廓，并使用类似 IoU 的损失函数优化渲染轮廓与分割之间的差异。在这一过程中，可微渲染器是梯度传播的媒介，其关键步骤被称作可微光栅化（Differentiable Rasterization），把图片的 Alpha channel 变成了一个受像素到 Mesh 中三角面片的距离影响的概率分布图。3DDFA-V3 简单讨论了“距离影响概率”这一可微渲染的基本设定可能存在一些问题：一方面，为了让产生的梯度有效地影响三维形变，被渲染的像素点应该受到全局的三角面片的纹理等属性的影响；另一方面，从视觉渲染效果角度出发，为了让渲染像素点的纹理足够清晰，只能让其尽可能只受到离它最近的三角面片的影响；两者存在矛盾，导致三角面片距离像素点较远时，几乎接受不到梯度影响，如图 3 所示。因此，3DDFA-V3 认为可微渲染器对形状约束的能力不强，希望利用分割信息设计更直接有效的损失函数引导三维人脸形变。

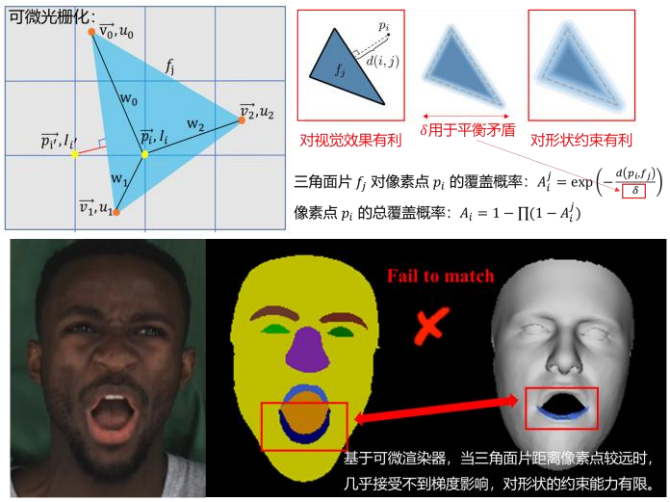


图 3　可微渲染器对形状约束能力不强的原因分析

3DDFA-V3 的关键思想是将目标和预测的部件分割转化为语义点集，通过优化点集的分布来确保重建区域和目标具有相同的几何形态。具体来讲，3DDFA-V3 提出了部件重投影距离损失(Part Re-projection Distance Loss, PRDL)。PRDL 按照区域P={left-eye, right-eye, left-eyebrow, right-eyebrow, up-lip, down-lip, nose, skin}对人脸进行分块，针对二维部件分割的每个部分$p \in P$，PRDL 首先在分割区域内采样点，得到目标点集$\{C_p \mid p \in P\}$。然后，PRDL 将三维人脸重建结果重新投影到图像平面上，并根据人脸模型的 masks 获得与目标区域$\{C_p \mid p \in P\}$语义一致的预测点集$\{V_{2d}^p(\alpha) \mid p \in P\}$，$\alpha$是人脸模型的系数。接着 PRDL 对图像平面的网格点进行采样，得到锚点集合A，并计算任意一个锚点$a_i \in A$到点集的各种统计距离（如最近距离f_{min}、最远距离f_{max}、平均距离f_{ave}等）来建立几何描述子。最后，PRDL 通过优化相同语义的预测点集的几何

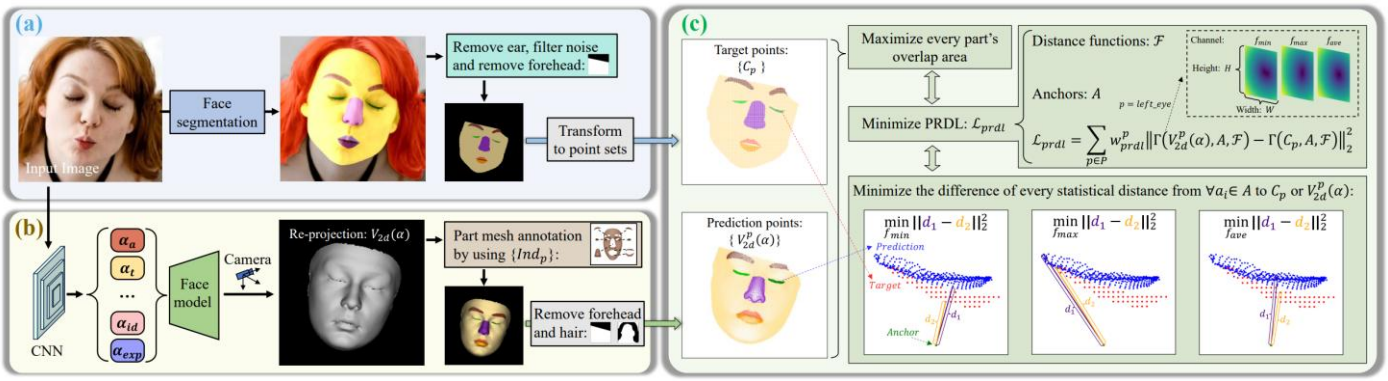


图 4　部件重投影距离损失（PRDL）概述，其核心思想是最小化每个锚点到目标和预测点集的统计距离的差异

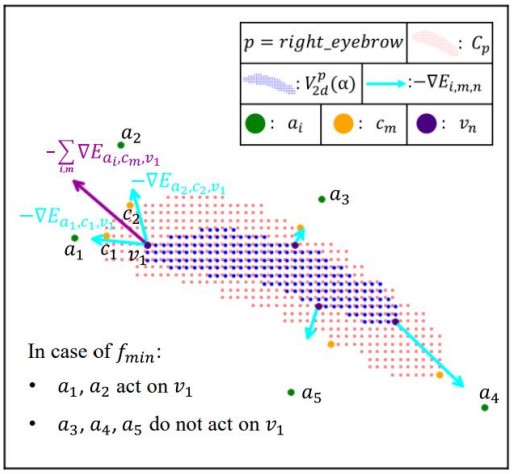


图 5　PRDL 的梯度分析图，PRDL 的锚点a_i对预测点有选择以及“推”和“拉”的作用，最终的作用在预测点上的梯度是相关锚点作用的叠加，具有鲁棒性。

描述子和目标点集的几何描述子的差异，确保重建区域和目标具有相同的几何分布，从而提高目标和预测点集覆盖区域之间的重叠度，整个过程如图 4 所示。概括来讲，对于$\forall p \in P$和$\forall a_i \in A$，PRDL 的优化目标为：

$$For\ \forall p \in P, \forall a_i \in A:$$
$$\begin{cases} \min \left\| f_{min}(V_{2d}^p(\alpha), a_i) - f_{min}(C_p, a_i) \right\|_2 \\ \min \left\| f_{max}(V_{2d}^p(\alpha), a_i) - f_{max}(C_p, a_i) \right\|_2 \\ \min \left\| f_{ave}(V_{2d}^p(\alpha), a_i) - f_{ave}(C_p, a_i) \right\|_2 \end{cases}$$

在使用过程中，PRDL 可以使用最远点采样(Farthest Point Sampling)等技术可以减少C_p、$V_{2d}^p(\alpha)$和A中的点数量，从而降低计算成本。通过理论推导可知，在梯度下降的条件下，PRDL 的锚点a_i对预测点有选择以及“推”和“拉”的作用，最终能使得锚点a_i到预测点集和目标点集的统计距离尽可能相同，从而指导三维人脸的形变，如图 5 所示，当预测点被目标点包围时，锚点可以对预测点产生向外扩展的梯度下降方向。

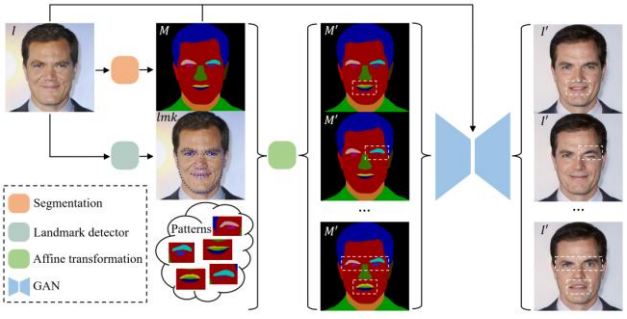


图 6　可控、大批量地合成闭眼、歪嘴、皱眉等表情

研究内容 2：数据策略的研究

虚拟合成数据常用于训练三维人脸重建模型。现有的合成人脸数据要么侧重于背景、光照和身份的多样化，要么集中在姿态变化上，虽然在重建自然面部表情方面提供了有效的帮助，但在重建极端表情方面难以提供支持。为了克服这些局限并促进相关研究，3DDFA-V3 采用了一种基于 GAN 的方法来大批量可控地合成难以搜集的人脸极端表情数据，包括闭眼、歪嘴和皱眉等表情。

如图 6 所示，我们首先使用人脸分割方法和关键点检测器分别获取原始图像I的分割图M和关键点lmk，并预设一些人脸表情的局部变化模板$Patterns$。利用关键点lmk的位置信息，对$Patterns$进行合适的仿射变换，将其应用到原始分割图M上，得到M'。随后将M'输入到一个条件 GAN 网络中，生成新的面部表情图像I'。目前 3DDFA-V3 已经生成了超过 50 万张的闭眼、歪嘴、皱眉等表情的数据，并进行了开源，数据示例如图 7 所示。

2.3 实验分析与结果

详细的实验设置、定量或定性实验可参考 3DDFA-V3 论文，下面仅对关键实验进行展示。

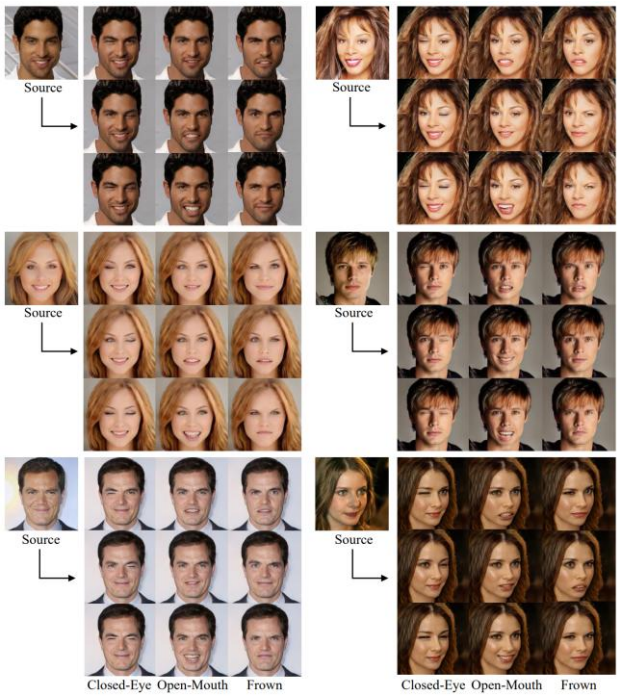


图 7　3DDFA-V3 提供的虚拟合成表情数据示例

Methods	*Frontal-view (mm)*↓					*Side-view (mm)*↓				
	Nose *avg.± std.*	*Mouth* *avg.± std.*	*Forehead* *avg.± std.*	*Cheek* *avg.± std.*	*avg.*	*Nose* *avg.± std.*	*Mouth* *avg.± std.*	*Forehead* *avg.± std.*	*Cheek* *avg.± std.*	*avg.*
PRNet[8]	1.923±0.518	1.838±0.637	2.429±0.588	1.863±0.698	2.013	1.868±0.510	1.856±0.607	2.445±0.570	1.960±0.731	2.032
MGCNet[9]	1.771±0.380	1.417±0.409	2.268±0.503	1.639±0.650	1.774	1.827±0.383	1.409±0.418	2.248±0.508	1.665±0.644	1.787
Deep3D[6]	1.719±0.354	1.368±0.439	2.015±0.449	1.528±0.501	1.657	1.749±0.343	1.411±0.395	2.074±0.486	1.528±0.517	1.691
3DDFA-v2[4]	1.903±0.517	1.597±0.478	2.447±0.647	1.757±0.642	1.926	1.883±0.499	1.642±0.501	2.465±0.622	1.781±0.636	1.943
HRN[7]	1.722±0.330	1.357±0.523	1.995±0.476	**1.072±0.333**	1.537	1.642±0.310	1.285±0.528	1.906±0.479	**1.038±0.322**	1.468
DECA[5]	1.694±0.355	2.516±0.839	2.394±0.576	1.479±0.535	2.010	1.903±1.050	2.472±1.079	2.423±0.720	1.630±1.135	2.107
Ours	**1.586±0.306**	**1.238±0.373**	**1.810±0.394**	1.111±0.327	**1.436**	**1.623±0.313**	**1.205±0.366**	**1.864±0.424**	1.076±0.315	**1.442**

表 1　3DDFA-V3 在 REALY benchmark 上取得了 SOTA 的水平

定量对比实验：如表 1 所示，3DDFA-V3 在 REALY benchmark 上取得了 SOTA 的水平。

定性对比实验：如图 8 所示，与现有的 SOTA 方法相比，3DDFA-V3 的预测结果可以非常准确地重建不对称和奇怪的面部表情。

三、基于素描草图的带纹理三维人脸重建

3.1 背景介绍

从人脸素描草图中重建带有纹理的精细 3D 人脸在刑侦与失踪人员调查、动漫娱乐、艺术设计等多个场景中具有广泛的应用潜力，是一个极具前景但尚未充分发

图 8　与现有的 SOTA 方法相比，3DDFA-V3 的预测结果可以非常准确地重建不对称和奇怪的面部表情

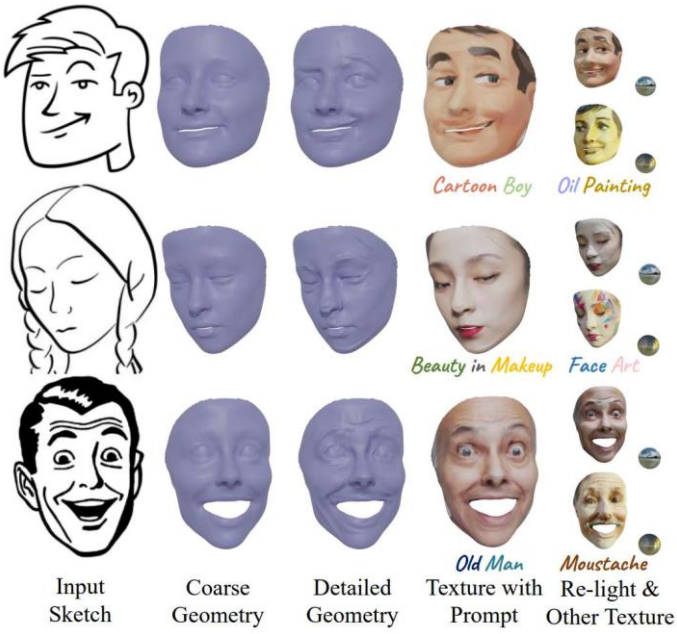


图 9 S2TD-Face 能够从不同风格的人脸草图中重建出高保真且拓扑一致的 3D 精细人脸

展的研究领域。现有研究主要面临着两方面的不足：一方面，现有的方法只能处理姿态受限且有真实阴影的人脸素描草图，且难以将素描的细节笔触反馈到重建的 3D 形状上；另一方面，纹理在面部外观的表现中起着关键作用，但素描草图缺乏这一信息，因此在重建过程中需要额外的纹理控制。

如图 9 所示，我们提出了一种从素描草图中重建具有可控纹理的 3D 人脸的新方法，称为 S2TD-Face（**S**ketch to controllable **T**extured and **D**etailed **T**hree-**D**imensional **F**ace）。S2TD-Face 引入了一个两阶段形状重建框架，能够直接从输入草图中重建精细的带纹理的三维人脸形状。为了将素描的细节笔触反馈到重建的 3D 形状上，S2TD-Face 提出了一种新的草图到几何形状的损失函数，以确保重建结果精确匹配输入特征，如草图勾勒出的酒窝和皱纹等。S2TD-Face 的训练不依赖难以获取的 3D 人脸扫描数据或手绘素描草图。此外，S2TD-Face 还引入了一个纹理控制模块，通过文本提示从纹理库中选择合适的纹理并将其无缝整合到几何结构中，从而得到具有可控纹理的 3D 细节人脸，此外，S2TD-Face 还支持基于 ControlNet 的 3D 人脸纹理控制方法。

3.2 研究内容

图 10 是 S2TD-Face 的总体流程概括，S2TD-Face 主要包括几何形状重建模块和纹理控制模块。

研究内容 1：训练策略的研究

基于现有大量的二维真实人脸图片，S2TD-Face 首先集成了各种素描草图生成方法，从二维人脸图片中得

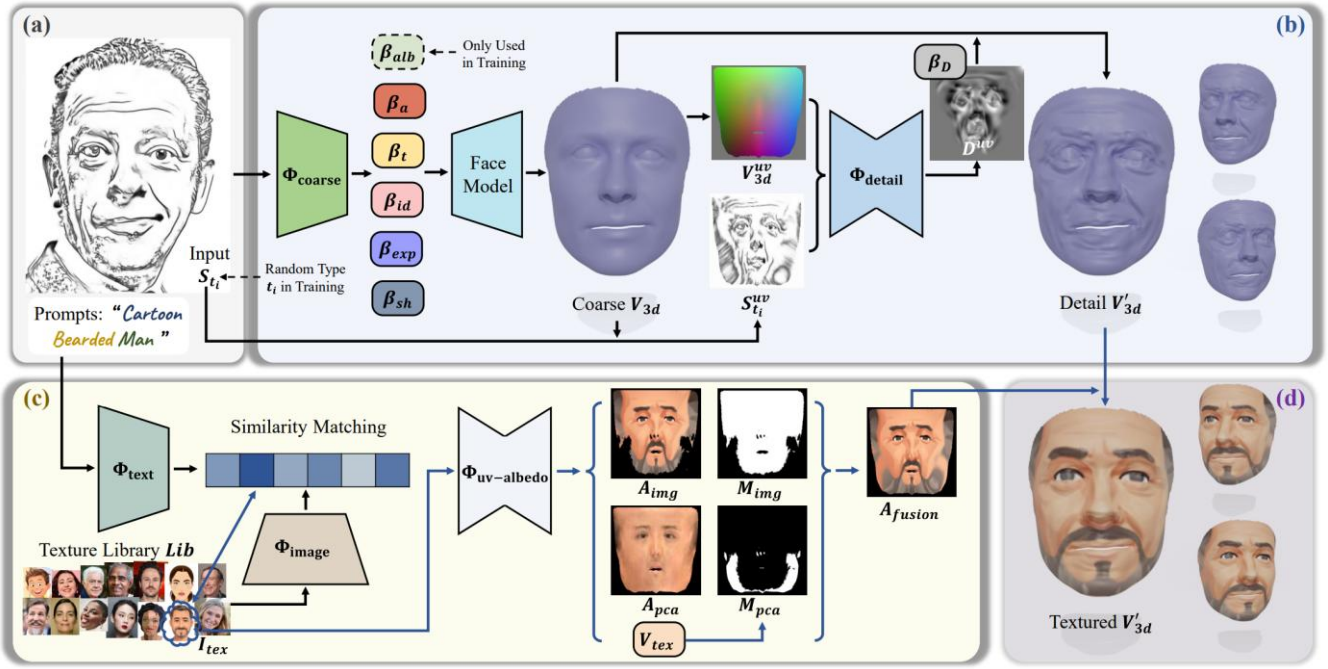


图 10　基于素描输入的纹理可控的三维精细人脸重建方法的流程概括

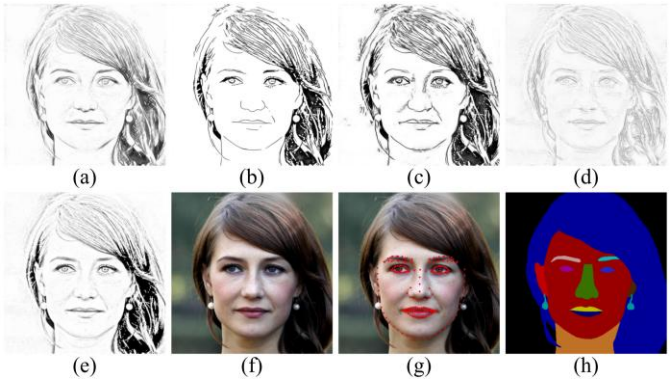


图 11　S2TD-Face 的数据示例

到了各类风格的素描人脸。由于每张素描人脸数据都有真实的二维人脸图片数据与之对应，S2TD-Face 在训练的时候能够结合成熟的三维人脸重建技术，利用已有的关键点信息、五官分割信息和纹理信息对重建出的 3D 形状进行约束。这种训练策略使得 S2TD-Face 不依赖于难以收集的 3D 人脸扫描数据和手绘草图。图 11 是 S2TD-Face 的数据示例，(a)-(e)为从原始图像(f)中生成的不同风格的草图，(g)表示关键点，(h)表示分割信息。S2TD-Face 重建框架的输入包括素描草图 (a)-(e)，(f)-(h)用作监督信号。

研究内容 2：重建与素描细节笔触一致的精细三维人脸

基于素描草图的特点，S2TD-Face 结合可微渲染技术，设计了有效的损失函数，其能够捕捉素描草图刻画的人脸形状信息，并将其准确地重建到三维结构上。S2TD-Face 使用法线偏移对人脸 mesh 进行精细化建模，并构建素描到三维信息的损失函数$\mathcal{L}_{\text{sketch}}$：

$$\mathcal{L}_{\text{sketch}} = \lambda_1 \underbrace{\sum_{n\in\{a,b,c,d\}} \|M^n - M\|_2}_{\text{sketch-photometric}} + \lambda_2 \underbrace{\sum_{n\in\{a,b,c,d\}} \left(1 - \frac{<\Phi_{\text{per}}(M^n), \Phi_{\text{per}}(M)>}{\|\Phi_{\text{per}}(M^n)\|_2 \cdot \|\Phi_{\text{per}}(M)\|_2}\right)}_{\text{sketch-perception}}$$

其中，M^n和M分别表示预测的素描图与对应真值素描图的经过面部区域掩码过滤的结果，$n \in \{a,b,c,d\}$表示预测素描图的四种形式，即a是由纹理和粗糙形状渲染得到的预测素描图；b是由纹理和精细形状渲染得到的预测素描图；c是由灰色纹理和粗糙形状渲染得到的预测素描图；d是由灰色纹理和精细形状渲染得到的预测素描图。$\mathcal{L}_{\text{sketch}}$包括两个部分，第一部分计算渲染预测的素描与对应真值的图片度量损失（sketch-photometric），第二部分计算渲染预测的素描与对应真值的感知度量损失 (sketch-perception)，$\mathcal{L}_{\text{sketch}}$的可视化过程如图 12 所示，更详细的计算过程也可参考 S2TD-Face 原文。

研究内容 3：精细三维人脸纹理控制模块

对于精细三维人脸纹理控制模块，S2TD-Face 首先搜集一定数量的各种外观风格的人脸图像，作为预设的人脸纹理模板库 Library。S2TD-Face 将使用者提供的待重建三维人脸的纹理的文本描述 Text，作为纹理控制模块的输入。利用 CLIP 对预设的人脸纹理模板库 Library 中的图片进行匹配，选取最相似的人脸纹理图片或从最相似的前 k 张图片中选取任意一张图片，保证方法的灵活性，利用三维人脸重建技术估计人脸图片的 UV 纹理展开图，并同时估计出 PCA 纹理对不可见区域进行补全。此外，最新开源的 S2TD-Face 还进一步支持了基于 ControlNet 的 3D 人脸纹理控制方法。

3.3 实验分析与结果

详细的实验设置、定量或定性实验可参考 S2TD-Face 论文，下面仅对关键实验进行展示。

可视化结果：如图 13 所示，S2TD-Face 能够从不

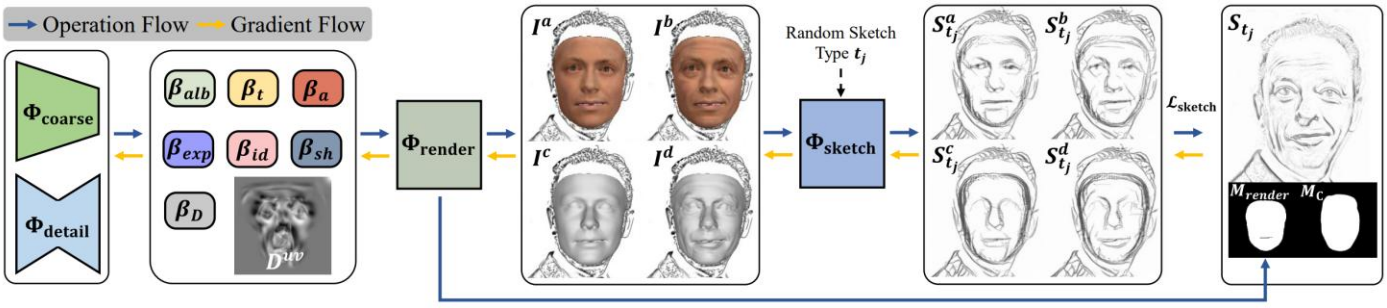


图 12　素描到三维信息的损失函数$\mathcal{L}_{\text{sketch}}$的过程概述

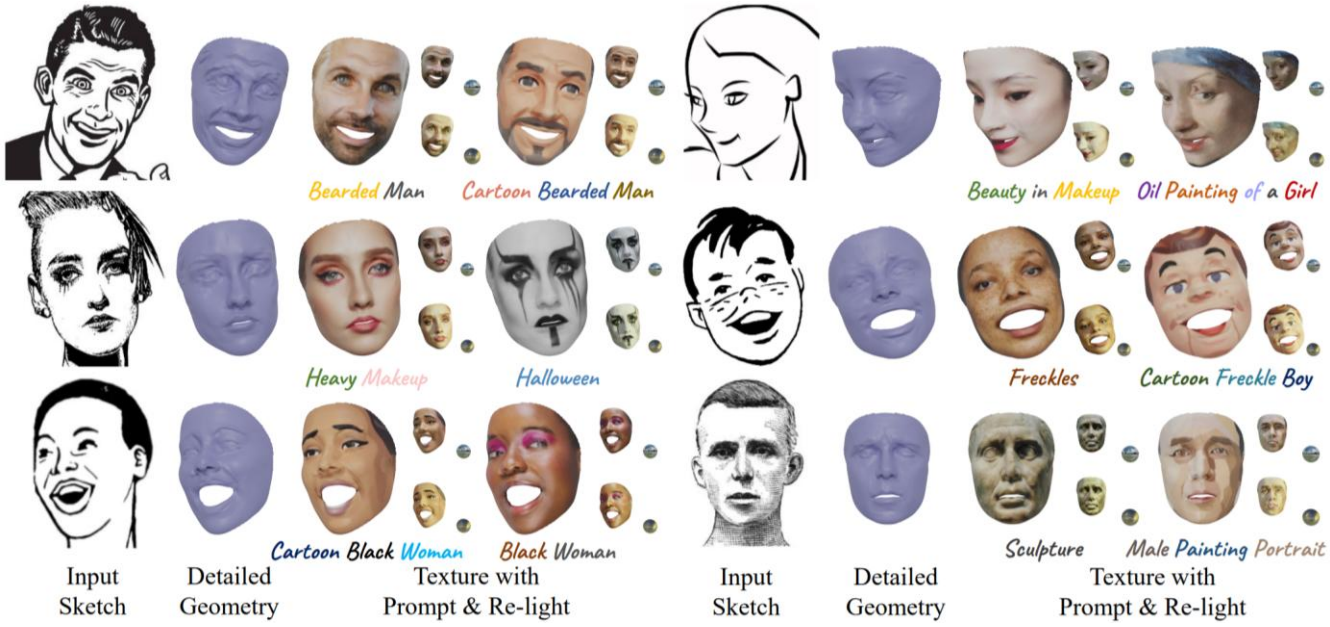


图 13　S2TD-Face 的可视化结果

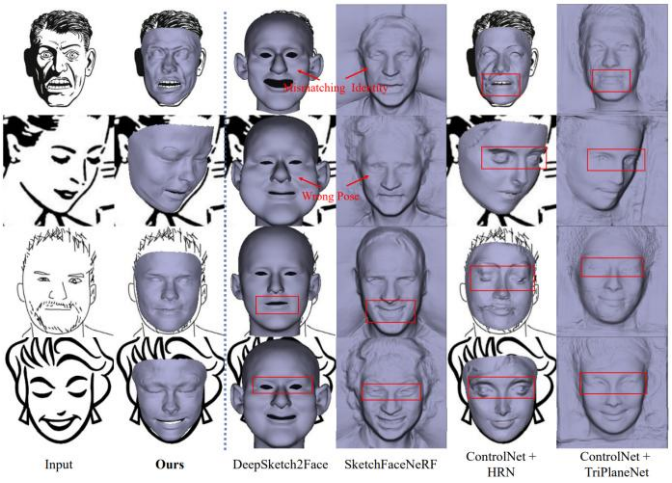


图 14　S2TD-Face 与现有方法对比

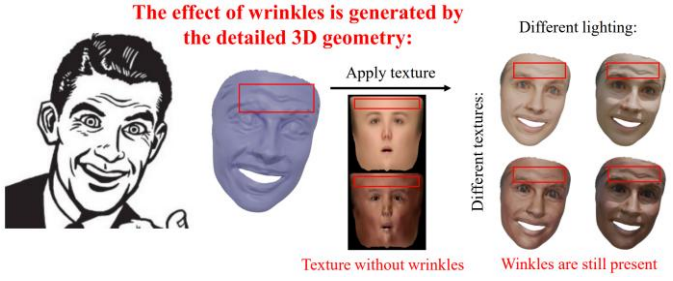


图 15　S2TD-Face 的局部细节效果（如皱纹、酒窝等）是由形状表示的，不依赖于特定的纹理

同风格的人脸草图中重建出高保真且拓扑一致的 3D 精细人脸。它还支持基于文本提示的 3D 人脸纹理控制，能够生成卡通、雕塑风格或真实人脸风格的纹理。

对比实验：如图 14 所示，与现有方法相比，S2TD-Face 的重建结果展现出了与输入人脸素描草图细节和身份高度一致的最佳效果。

3D 形状细节的捕捉方式：如图 15 所示，得益于有效的几何形状重建模块和纹理控制模块的独立设计，S2TD-Face 对人脸素描草图刻画的细节的捕捉方式是基于 3D 精细形状的，不受纹理变化的影响。

四、总结展望

三维人脸精细结构重建作为三维人脸领域的重要方向，对于数字人驱动、精神状态分析等实际应用具有深远意义。尽管现有方法在精细化建模方面取得了显著进展，但仍存在局部细节捕捉不足、动态表情重建受限等挑战，此外，与可微渲染流程结合设计更有力的形状约束方式也是值得探讨的问题。本文以三维人脸精细结构重建为主题，讨论了基于部件分割引导的人脸细节捕捉（3DDFA-V3）和基于素描草图的带纹理三维人脸重建（S2TD-Face）这两项科研工作，希望为该领域的进一步发展提供借鉴和启发。本文介绍的 3DDFA-V3 和 S2TD-Face 的代码请参考 https://github.com/wang-zidu/3DDFA-V3 和 https://github.com/wang-zidu/S2TD-Face。

责任编委 张 杰

参考文献

[1] Wang, Zidu, et al. 3D Face Reconstruction with the Geometric Guidance of Facial Part Segmentation. In CVPR 2024.

[2] Wang, Zidu, et al. S2TD-Face: Reconstruct a Detailed 3D Face with Controllable Texture from a Single Sketch. In ACM MM 2024.

[3] Zhu, Xiangyu, et al. Face alignment in full pose range: A 3d total solution. In IEEE T-PAMI 2017.

[4] Guo, Jianzhu, et al. Towards fast, accurate and stable 3d dense face alignment. In ECCV 2020.

[5] Feng, Yao, et al. Learning an animatable detailed 3D face model from in-the-wild images. In ToG 2021.

[6] Deng, Yu, et al. Accurate 3d face reconstruction with weakly-supervised learning: From single image to image set. In CVPR workshops 2019.

[7] Lei, Biwen, et al. A hierarchical representation network for accurate and detailed face reconstruction from in-the-wild images. In CVPR 2023.

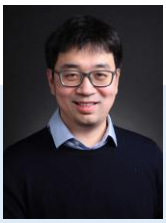
[8] Feng, Yao, et al. Joint 3d face reconstruction and dense alignment with position map regression network. In ECCV 2018.

[9] Shang, Jiaxiang, et al. Self-supervised monocular 3d face reconstruction by occlusion-aware multi-view geometry.

王子都

王子都，中国科学院自动化研究所 2022 级硕士研究生，导师为朱翔昱副研究员，主要研究方向为三维人脸重建。

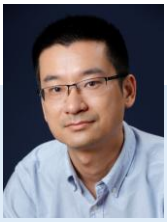
Email: wangzidu2022@ia.ac.cn



朱翔昱

朱翔昱，中国科学院自动化研究所副研究员，从事生物特征识别、数字人、三维重建等方向的理论研究与应用。国际模式识别协会（IAPR）生物特征青年学者奖（YBIA）获得者（两年一次，每次从全球范围内评选 40 岁以下学者一名）。任 CCF:A 类期刊 T-IFS Associate Editor。共发表论文 80 余篇，发表文章的 Google Scholar 总引用次数为 9600 余次。获得三次国际竞赛冠军以及四项最佳论文及提名奖。授权国家发明专利 12 项。入选 IEEE Senior Member，百度学术全球华人 AI 青年学者榜单（全球 25 人），受到腾讯犀牛鸟基金支持。获 2021 中国电子学会科技进步二等奖、中国图象图形学学会优秀博士论文提名奖。提出的人脸三维建模方法在被 PyTorch 官方 Twitter 报道，开源代码在 GitHub 上收获 6000 余星。

Email: xiangyu.zhu@nlpr.ia.ac.cn



雷 震

雷震，IEEE/IAPR Fellow，中国科学院自动化研究所研究员，中国科学院大学岗位教授，中国科学院香港创新院人工智能与机器人创新中心教授，博士生导师。致力于人工智能基础理论，图像视频分析与理解，生物特征识别方面的研究。Google Scholar 文章引用次数 3.3 万余次，H-index：85，2020-2023 爱思唯尔中国高被引学者，授权发明专利 30 余项，撰写发布国家标准 2 项，国家公共安全行业标准 7 项，十余次获得国际学术会议最佳（学生）论文奖或国际会议竞赛第一名。

Email: zhen.lei@ia.ac.cn

热点追踪

基于三维高斯的动态人脸重建与几何纹理联合优化的语音驱动人脸

上海交通大学 李炫辰 程宇豪 晏轶超

本文是上海交通大学人工智能研究院晏轶超副教授团队的系列研究成果。高精度动态三维人脸资产的重建与驱动是三维数字人产业的核心问题，长期受到学术界和产业界的广泛关注。在动态人脸重建方面，当前业界广泛使用多目立体视觉和非刚性配准结合的方法。然而这种方法容易失效，且需要专业艺术家耗费大量时间精力进行手工修饰。针对这一痛点问题，团队提出了首个基于拓扑一致的三维高斯人脸表示的动态人脸重建框架Topo4D[1]，能够高效重建动态高保真人脸几何和8K分辨率纹理贴图，加速传统方法数十倍，且具有优秀的时序稳定性和面对极端表情的鲁棒性（如图1）。在语音驱动人脸方面，现有的工作都只聚焦于单独驱动人脸几何，而忽视了随几何一致变化的动态纹理贴图的生成。为了推动人脸几何与纹理的协同驱动生成研究，团队基于Topo4D的高效重建能力，兼顾效率与精度，构建了一个具有100个受试者的扫描级语音-模型-贴图对齐数据集TexTalk4D。基于该数据集，团队提出了首个基于扩散的语音协同驱动几何纹理框架TexTalker，能够根据任意语音输入同时生成与音频对齐的动态人脸模型和与面部运动一致变化的动态纹理贴图。大量的实验证明该方法在几何纹理精度和一致性方面都优于已有方法。论文[1]的项目网页公开在https://xuanchenli.github.io/Topo4D/.

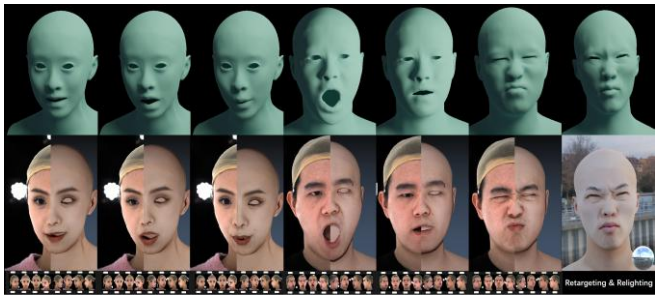


图 1　Topo4D 的结果示例。Topo4D 能够生成时序一致、拓扑一致的高保真网格模型和 8K 分辨率贴图。重建的资产可以直接应用于重定向和重打光。

一、研究背景与相关工作

1.1 高精度面部资产重建

获取具有预定义拓扑结构的高保真面部模型和贴图是业界一个长期存在的研究难点。随着3D可变形人脸被提出，许多人脸重建方法利用参数化模型实现基于单图或多目图像的面部资产重建。然而受限于参数化模型的表现力，这些方法难以忠实重建极端表情。在工业界常用非刚性ICP方法配准由多目立体视觉重建的高模。然而直接将该方法扩展到动态人脸的重建会导致时序上不稳定，导致纹理漂移等问题。为了实现稳定的动态重建，现有CG管线通常使用专业软件逐帧对高精度模型进行重拓扑，涉及经验丰富的艺术家投入大量时间和专业知识。为了解决这一问题，一些方法利用光流或其他技术作为监督来变形模板网格。然而，这些方法需要细致调整大量超参数，难以泛化到不同的个体，并且计算速度较慢。当前也有方法直接从图像中回归人脸模型，但这些方法需要大量昂贵的训练数据，并且难以扩展到新的采集系统，同时对于数据集之外的表情也容易失效。本文中我们提出的Topo4D能够在显著降低时间和人力成本的同时，获取通常需要艺术家手动制造的高质量面部模型和贴图。

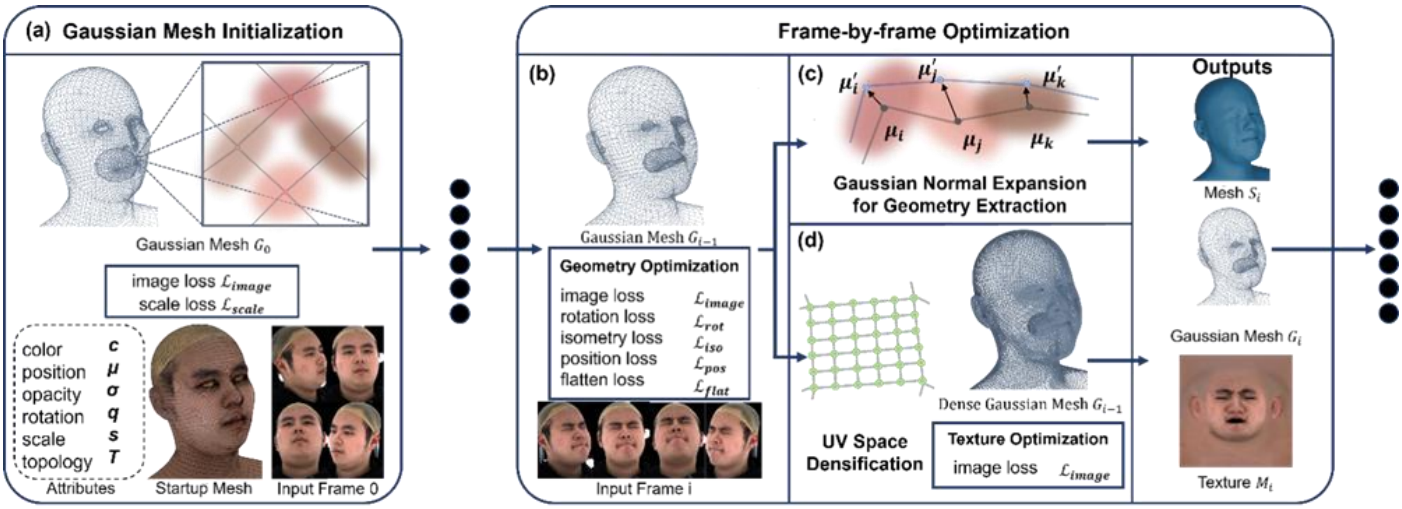


图 2　Topo4D 框架示意图。（a）初始化高斯网格并构建高斯点之间的拓扑关联。（b）利用物理先验与几何先验损失优化几何。（c）通过高斯法向偏移拉近高斯表面与渲染表面。（d）执行 UV 空间稠密化学习细致毛孔级纹理细节。

1.2 语音驱动三维人脸动画

语音驱动三维人脸动画的主要难点在于学习音素与视位素之间的复杂映射。传统的基于规则或数据驱动的方法利用手工设计的映射关系从音节中生成网格变形，该方法人力成本高，且无法泛化到新的语言。许多基于学习的方法能够从对齐数据中挖掘音频特征与人脸运动之间的关联。虽然这些工作已经能够生成生动的面部运动，却都忽视了动态纹理贴图的生成。如果没有动态纹理，则无法表现人脸运动过程中的毛孔挤压和褶皱变化，将会大大降低渲染真实感，甚至导致恐怖谷效应。

为了获取与面部肌肉运动一致变化的纹理贴图，传统CG管线通常从个性化的高精度表情基底中构建专属的动态纹理库，并根据顶点位移采样纹理库来实现纹理变化。该方法需要高成本的表情资产，无法泛化到不同的人，同时计算开销巨大。当前已有方法能够生成动态纹理，但是都忽视了时序稳定性。为了更好地研究几何与纹理之间的一致关联学习，促进几何与纹理的协同生成研究，我们提出了一个新的扫描级语音-模型-贴图对齐数据集TexTalk4D，包含100个个体的说话数据。我们基于此数据集提出的几何纹理协同生成框架TexTalker通过统一纹理与几何的表示，利用隐扩散模型在低维隐空间中学习几何与纹理的复杂关联，实现了高度一致和稳定的语音驱动人脸动画生成。

二、Topo4D

本节介绍提出的Topo4D，整体框架如图2所示。该方法旨在从多视角视频中实现时序稳定的头部重建和纹理生成。具体来说，给定来自K个视角的共F帧分辨率为$h \times w$的多视角图像序列$\{I_i^j \in \mathbb{R}^{h\times w\times 3} | 0 \le i \le F-1\}$，我们的方法能够重建具有固定拓扑T的头部网格模型序列$\{S_i := (V_i, T) | V_i \in \mathbb{R}^{n_v\times 3}\}_{i=0}^{F-1}$和8K纹理贴图序列$\{M_i \in \mathbb{R}^{8192\times 8192\times 3}\}_{i=0}^{F-1}$，其中$n_v$为顶点个数。

接下来本节分别介绍Topo4D中的高斯网格初始化、动态几何与纹理交替优化、以及从高斯网格中提取高质量资产的方法。最后我们通过展示实验结果证明Topo4D的优越性。

2.1 高斯网格初始化

原始三维高斯场景由于缺少拓扑关系，难以直接从中提取可用的网格模型。为了解决这个难点，我们提出了高斯网格人脸来将网格拓扑关系结合到三维高斯表示中。具体来说，我们将第i帧的人脸表示为一个三维高斯点集$G_i = \{G_{i,j}\}_{j=1}^{n_v}$，其中每个高斯点都有一组可学习参数：颜色（$\boldsymbol{c}$）、坐标（$\boldsymbol{\mu}$）、透明度（$\boldsymbol{\sigma}$）、旋转四元数（$\boldsymbol{q}$）、尺寸（$\boldsymbol{s}$）。为了获取预定义的拓扑结构，我们利用一个初始网格模型的顶点坐标和贴图颜色来初始化第一帧的高斯网格。并利用图像差异损失（$\mathcal{L}_{image}$）和大小约束损失($\mathcal{L}_{scale}$)在多视角图像上优化初始高斯网格的各个属性来更好地用高斯网格人脸拟合真实人脸：

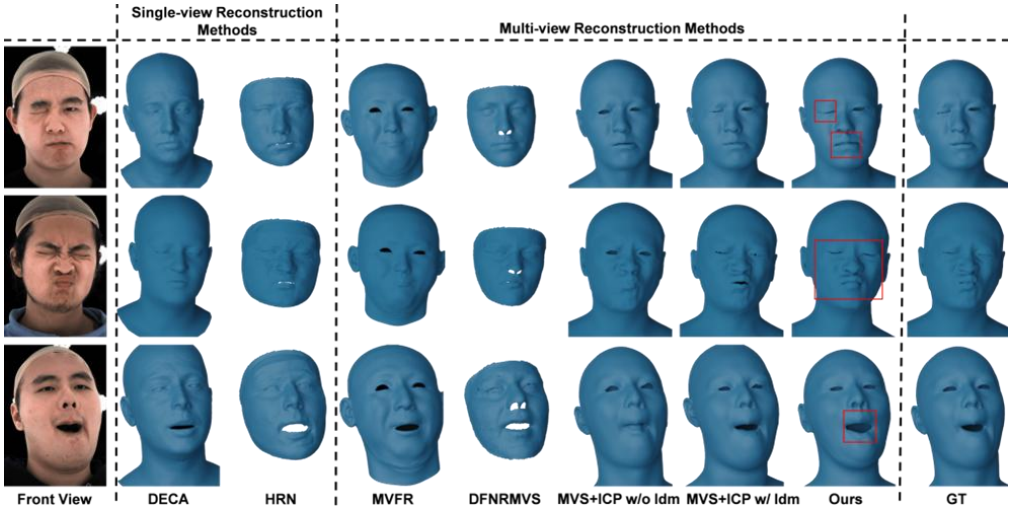


图 3　Topo4D 和其他拓扑一致的重建方法的重建结果比较。我们使用艺术家手工配准的网格作为 GT，红框突出了难以重建的区域。

$$\mathcal{L}_{\text{image}} = (1-\lambda_{image})\mathcal{L}_1(\mathbf{I}_0, \mathbf{I}_0') + \lambda_{image}\mathcal{L}_{D-SSIM}(\mathbf{I}_0, \mathbf{I}_0'),$$

$$\mathcal{L}_{\text{scale}} = \sum_{i\in G}(\|\boldsymbol{s}_{0,i}\|_{-\infty} + \boldsymbol{max}(0, \boldsymbol{s}_{0,i} - \lambda_{init}\boldsymbol{s}_{init,i})),$$

其中图像差异损失比较真实照片$\mathbf{I}_0$和渲染结果$\mathbf{I}_0'$之间的差异，用$\mathcal{L}_1$损失和$\mathcal{L}_{D-SSIM}$损失加权表示。大小约束损失用于约束高斯点的尺寸，目的是限制高斯点法向上的尺寸接近于0，同时整体尺寸不超出初始值太多，使得高斯分布之间的重叠尽可能少，同时让高斯点贴合人脸表面。

2.2 几何纹理交替优化

在初始化得到第一帧的高斯网格后，我们提出了一种交替几何和纹理优化方法来在前一帧的基础上优化得到当前帧的高斯网格几何，并从多视角图像中学习细节的纹理颜色。

几何优化 直接优化高斯点的位置属性会导致模型顶点错乱。为了在优化过程中维持高斯点之间的拓扑结构和规则的网格排列，我们引入了基于物理先验和基于拓扑先验的损失函数来约束优化过程。该阶段损失包含三项：$\mathcal{L}_{\text{geo}} = \mathcal{L}_{\text{image}} + \mathcal{L}_{\text{phy}} + \mathcal{L}_{\text{topo}}$.

具体来说，面部的低频细节区域会导致高斯点跟踪错误，从而导致破坏网格的拓扑结构。在Luiten[2]等人提出的物理约束的基础上，我们向其中引入了拓扑关系，让每个高斯点受到其一环邻域中的点的约束：

$$\mathcal{L}_{\text{rot}} = \frac{1}{2n_e}\sum_{i\in G}\sum_{j\in K_i}\omega_{i,j}\|\hat{q}_{t,j}\hat{q}_{t-1,j}^{-1} - \hat{q}_{t,i}\hat{q}_{t-1,i}^{-1}\|_2,$$

其中$\hat{q}$为归一化四元数，n_e为边数，ω为影响权重，K_i为高斯点$G_{t,i}$的邻接点。影响权重考虑了初始边长：

$$\omega_{i,j} = exp(-\lambda_\omega\|\mu_{0,j} - \mu_{0,i}\|_2^2).$$

除了对约束高斯点的旋转相似性，我们发现对点于点之间距离施加刚性约束有助于保持长时间范围内的稳定高斯点跟踪：

$$\mathcal{L}_{\text{iso}} = \frac{1}{2n_e}\sum_{i\in G}\sum_{j\in K_i}\omega_{i,j}\left|\|\mu_{0,j} - \mu_{0,i}\|_2 - \|\mu_{t,j} - \mu_{t,i}\|_2\right|.$$

总之，物理先验损失为上述两项的加权组合：

$$\mathcal{L}_{\text{phy}} = \lambda_{rot}\mathcal{L}_{\text{rot}} + \lambda_{iso}\mathcal{L}_{\text{iso}}.$$

物理先验损失保证了高斯点动态跟踪的稳定性，但并不保证高斯网格表面和布线的规整。为了解决这个问题，我们进一步引入拓扑先验损失。具体来说，我们让每个高斯顶点向其邻接点的中心位置靠近：

$$\mathcal{L}_{\text{pos}} = \frac{1}{n_v}\sum_{i\in G}(\mu_{t,i} - \frac{\sum_{j\epsilon K_i}\mu_{t,j}}{|K_i|})^2.$$

Type	Methods	<0.2mm (%)↑	<0.5mm (%)↑	<1mm (%)↑	<2mm (%)↑	<3mm (%)↑	Mean (mm)↓	Med. (mm)↓
Single-view	DECA	2.055	5.136	10.264	20.075	29.046	8.104	5.929
	HRN	5.170	12.786	20.734	44.692	60.263	2.871	2.429
Multi-view	MVFR	4.139	10.130	19.407	34.629	43.661	7.800	4.357
	DFNRMVS	3.447	8.579	17.064	33.479	48.356	3.649	3.214
	Topo4D (Ours)	**22.485**	**52.856**	**87.376**	**94.379**	**97.697**	**0.686**	**0.471**

表 1　不同的人脸重建方法的定量比较。我们计算了在不同的误差水平内的顶点所占的百分比，并计算平均误差和中位数

为了获取平滑的表面，防止模型出现穿模和突刺，我们同时对面片之间的夹角进行约束：

$$\mathcal{L}_{\text{flat}} = \sum_{\theta_i \epsilon e_i}(1 - cos(\theta_{t,i} - \theta_{0,i})),$$

其中$\theta_{t,i}$为第t帧公共边为e_i的面片夹角。

总之，我们的拓扑先验损失为上述两项的加权组合：

$$\mathcal{L}_{\text{topo}} = \lambda_{\text{pos}}\mathcal{L}_{\text{pos}} + \lambda_{\text{flat}}\mathcal{L}_{\text{flat}}.$$

纹理优化 在获得当前帧几何后，我们继续学习高细节的纹理。学习8K纹理需要大量的高斯点，与传统3DGS的致密化方法不同，我们提出UV空间稠密化。具体来说，在第t帧，我们首先用高斯网格G_t初始化稠密高斯网格G_t'。然后，我们通过双线性插值向每个四边形网格插入$N \times N$个高斯点，并初始化其属性。我们通过图像损失从4K原始图像中学习细节的高频纹理。

2.3 高质量资产提取

在当前帧优化结束后，可以从G_t和G_t'中提取网格和纹理。为了让高斯表面更加贴近渲染表面，我们通过高斯法向偏移让高斯中心偏移至高斯分布边缘，并直接从中提取顶点坐标。为了获取纹理，我们将G_t'中学习到的颜色属性通过预定义的UV坐标映射渲染至UV贴图空间。

2.4 相关实验

数据集 我们用光场相机采集了一个包含10个个体的数据集用于实验。每个受试者分别表演一段随机的表情和说一段话，包括各种极端和不对称的表情。每段视频长约10秒，每帧包含来自16个视角的分辨率为4096×3000的同步图像，帧率为60 fps。

几何精度评估 我们将Topo4D与三种拓扑一致的人脸重建方法进行比较：(1) 单目方法DECA[3]和HRN[4]；（2）多目方法MVFR[5]和DFNRMVS[6]；（3）使用人脸关键点引导的传统MVS和ICP结合方法。

图3展示了定性对比的结果，我们的方法可以忠实地捕捉不对称的极端表情和微小的面部变化，优于其他基于学习的方法。此外，我们将我们的方法与传统的基于优化的MVS+ICP管线进行了比较，同时在其中使用了关键点作为引导。即使是最先进的关键点检测器在极端表情下也会有较大的误差，导致配准错误，并且在精细的面部区域容易发生穿模。相比之下，我们的方法在不使用任何额外监督的情况下实现了正确的重建。

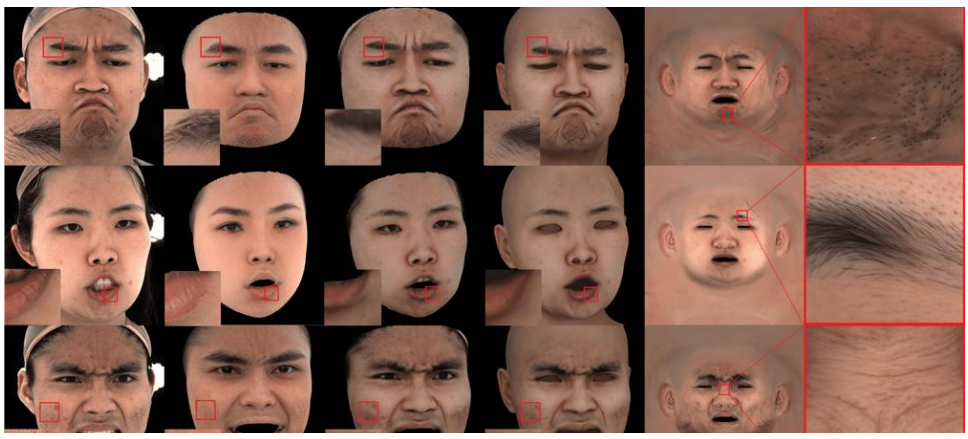
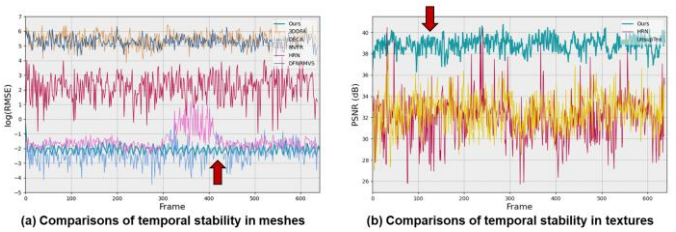


图 4　纹理质量比较。第 5 列和第 6 列展示了 Topo4D 生成的 8K 纹理贴图和其中的毛孔级细节



(a) Comparisons of temporal stability in meshes (b) Comparisons of temporal stability in textures

图 5　时间稳定性比较，Topo4D 用红色箭头标出

表1展示了定量比较的结果，我们计算了顶点到扫描模型表面的距离作为误差。我们的方法在所有指标上都显著优于其他拓扑一致的方法。由于我们在高斯初始化过程中引入了可靠的几何形状和颜色先验，大多数顶点都位于高精度范围内，在0.5mm精度范围内的顶点超过52.8%。

纹理质量评估 我们和最先进的面部纹理估计模型定性对比纹理质量，包括UnsupTex[7]和HRN[4]。图4显示了不同方法的渲染结果和Topo4D生成的8K纹理。UnsupTex和HRN的纹理缺乏真实细节且分辨率低。相比之下，我们的方法直接生成高质量的8K纹理，而无需进行上采样，忠实地捕捉面部皱纹、头发和毛孔，并实现了明显优越的渲染质量。

时序稳定性评估 我们比较了重建的网格和纹理的时间稳定性。在网格方面，我们通过计算相邻帧之间的RMSE来衡量稳定性。图5（a）展示了每种方法在一段视频上的重建曲线。Topo4D重建结果更加稳定，而其他方法有很大的波动。在某些情况下DECA的稳定性更好。这是因为DECA没有捕捉到一些极端的表情或微小的面部变化，因此倾向于保持网格不变。相反我们的方法可以忠实地捕捉到极端或细微的表情，同时保持稳定。

纹理方面，我们通过计算相邻帧之间贴图的PSNR来衡量稳定性。如图5（b）所示，我们的方法具有最高的平均值和最低的方差，表明其时序稳定。

消融实验 我们对方法中的核心设计进行了消融实验来评估其作用。图6展示了消融各个损失项、高斯法向偏移和改变输入视角数量的结果。物理先验损失有助于实现正确的密集对应，尤其嘴唇和鼻孔这些容易被遮挡且顶点密集的部位。虽然高斯尺寸损失似乎对几何精度的影响有限，但有助于学习毛孔级纹理细节和避免模糊，如图7所示。拓扑先验损失有效保证了不可见区域的稳定，同时避免了网格穿模和突刺问题。我们评估了输入视角的数量对重建的影响，如图6所示。即使只有一半的输入视图，我们的方法仍然可以产生有竞争力的结果，证明其也适用于只有较少视角的拍摄系统。然而，当视角数减少到4时，重建网格表现出明显的失真。

图7展示不同稠密化密度对纹理质量的影响。随着高斯点数量减少，纹理变得模糊并丢失细节。

三、TexTalk4D数据集构建

Topo4D能够兼顾效率和精度，高效重建扫描级动态人脸模型。为了实现具有动态纹理的语音驱动人脸动画模型，需要一个包含高质量纹理的音频-网格-纹理对齐数据集。虽然当前已经有许多可用的动态人脸数据集，但这些数据集大多不包含纹理贴图，或只提供低分辨率、稳定性差、缺少毛孔级细节的贴图，难以用于训练。因此我们基于Topo4D构建了一个包含8K分辨率动态纹理的扫描级精度数据集。

我们采集和构建数据的流程如图8所示。具体来说，

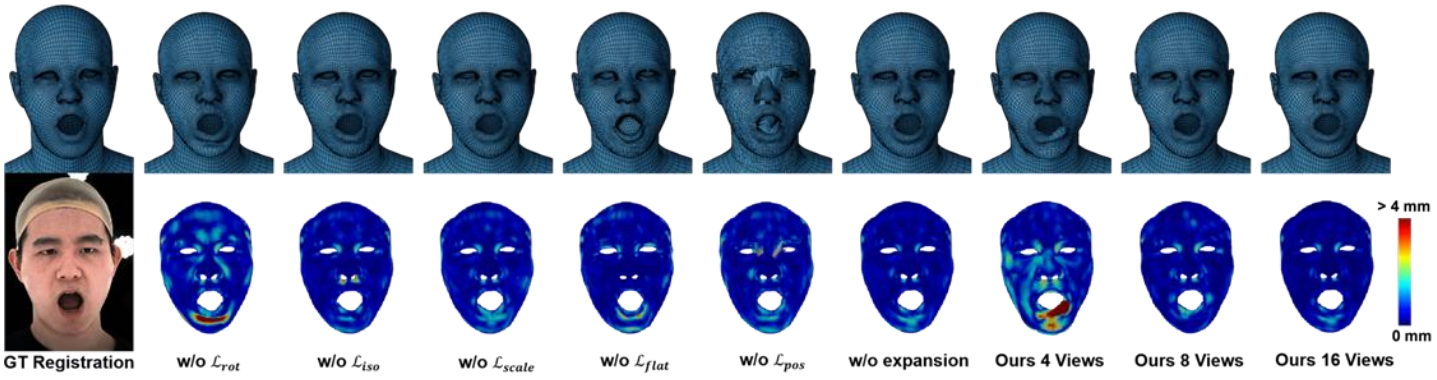


图 6　几何消融实验。第二行的热图展示了顶点距离真实表面的误差，颜色越亮代表误差越大。

我们手工重建了每段视频的第一帧模型作为初始化。由于原始的Topo4D难以重建高速的眨眼动作，我们通过引入人脸先验进行改进。在重建每一帧前，我们使用Mediapipe估计BS系数，并使用一套通用的表情基底作为当前帧的几何初始化。尽管估计的系数并不准确，但Topo4D可以在这个近似的基础上进行优化实现正确的重建。在获取了人脸几何后，我们将来自24个视角的4K图像颜色映射到UV空间中获取8K分辨率纹理贴图。我们通过将贴图与模板进行混合来消除头发等冗余区域。

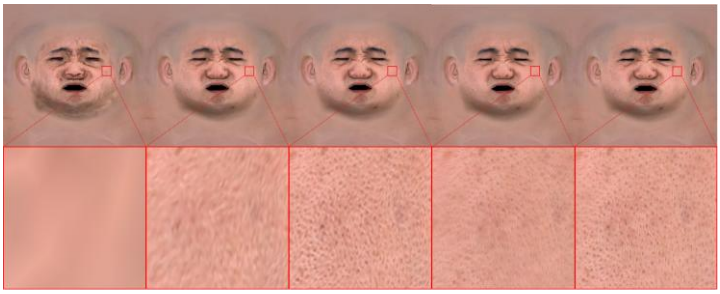


图 7　纹理消融实验

四、TexTalker

在TexTalk4D数据集的基础上，我们提出了一个能够音频同时驱动生成动态纹理与网格的模型TexTalker，如图9所示。我们的目标是从任意输入语音中生成具有个性化纹理的人脸动画资产。具体来说，给定T帧音频$A_{1:T}=(a_1,\dots,a_T)$，其中$a_t\in\mathbb{R}^D$的采样率为D，我们的方法能够生成面部运动序列$W_{1:T}=(w_1,\dots,w_T)$，和纹理序列$M_{1:T}=(m_1,\dots,m_T)$，其中$w_t\in\mathbb{R}^{n_f\times 3}$代表顶点位移，$m_t\in\mathbb{R}^{H\times W\times 3}$为用像素比值表示的纹理变化。除此之外，我们还希望能够控制生成的面部运动的说话风格和纹理变化的褶皱风格。

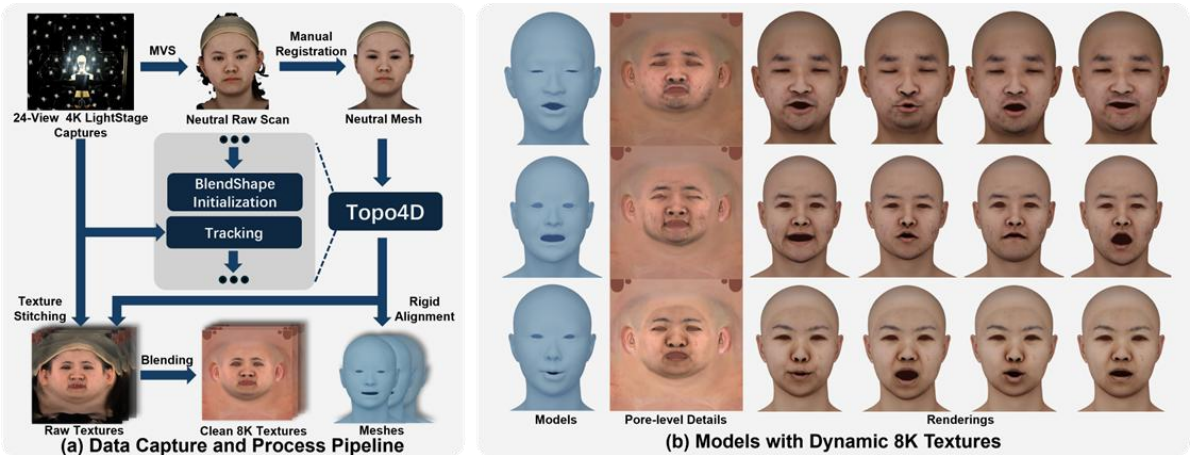


图 8　数据处理流程和数据集资产示例。我们使用 Topo4D 重建光场采集数据。在重建面部几何后，我们从 24 个视图的 4K 图像中映射贴图颜色来获取 8K 纹理。

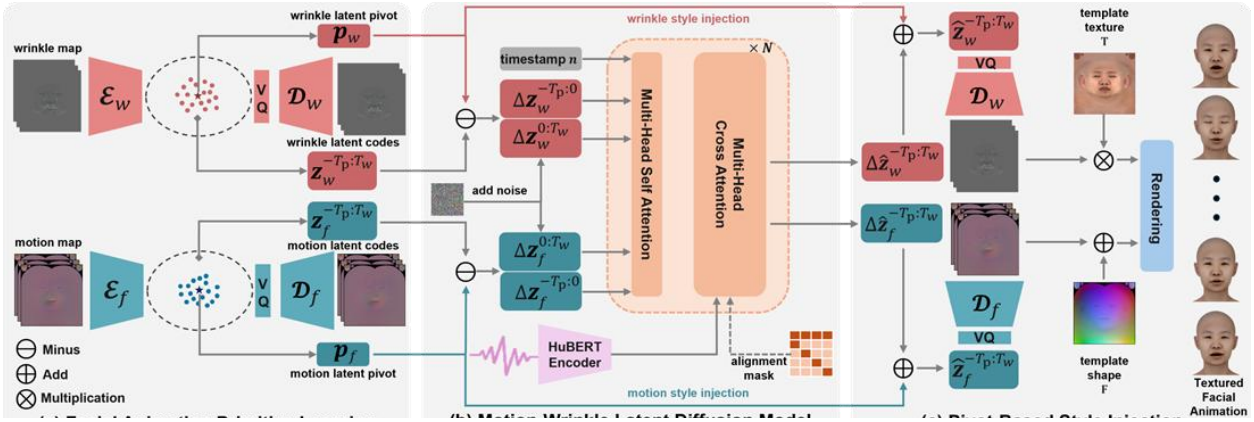


图 9　TexTalker 框架图。（a）我们训练了两个离散码本存储面部动画原语，并高效压缩贴图。（b）基于低维统一的几何与纹理表示，我们训练了隐扩散模型学习隐特征与均值特征的偏移。（c）通过将代表了风格的均值特征加回扩散模型的输出中，我们实现了解耦的说话和纹理风格控制。

为了实现这个任务，我们首先用相同的UV贴图空间统一表示面部运动和纹理变化，并在此基础上分别学习两个离散码本来存储面部动画原语并将贴图压缩至低维隐空间中。在此统一表征的基础上，我们训练了一个隐式扩散模型同时降噪生成几何和纹理隐式特征。最后，我们用隐空间中的平均特征来表示风格，并将其注入到扩散模型的输出中并解码得到所需的动画贴图，实现解耦的说话和纹理风格控制。整体框架如图9所示。

4.1 面部动画原语学习

为了统一几何与纹理的表示。在几何方面，我们通过UV映射将相对于无表情人脸的顶点位移映射到UV贴图空间得到面部运动贴图f。在纹理方面，我们使用和Zhang等人[8]同样的褶皱贴图w来表示纹理变化。

直接生成贴图会导致巨大的计算开销。受到离散先验表示的启发，我们通过训练离散自编码器将贴图压缩至低维空间中。我们分别训练一个几何自编码器$\mathcal{E}_f$和纹理自编码器$\mathcal{E}_w$来分别压缩几何运动和褶皱贴图，后续的生成模型也只需要以较低的开销学习生成隐特征即可。以几何自编码器为例，我们采用VQGAN[9]相同的方式联合训练编码器$\mathcal{E}_f$、码本$\mathcal{C}_f$、生成器$\mathcal{G}_f$、判别器$\mathcal{D}_f$。编码器将贴图压缩成低维隐特征$z_f = \mathcal{E}_f(f) \in \mathbb{R}^{h\times w\times d}$。通过对码本进行最近邻搜索，可以将将该隐特征变换为离散原语特征$\tilde{z}_f = \mathcal{Q}_f(z_f)$。最终，生成器可以从离散原语特征中重建出原始贴图$\hat{f} = \mathcal{G}_f(\tilde{z}_f)$。纹理自编码器以相同的方式实现压缩和重建。训练损失包括 (1) $\mathcal{L}_1$重建损失$\mathcal{L}_{\text{rec}}$，(2) VGG感知损失$\mathcal{L}_{\text{per}}$，(3) 对抗损失$\mathcal{L}_{\text{adv}}$，(4) 码本损失$\mathcal{L}_{\text{code}}$。总损失为以上损失项的加权组合：

$$\mathcal{L}_{\text{latent}} = \mathcal{L}_{\text{rec}} + \eta_{per}\mathcal{L}_{\text{per}} + \eta_{adv}\mathcal{L}_{\text{adv}} + \eta_{code}\mathcal{L}_{\text{code}}.$$

4.2 几何纹理协同优化隐式扩散模型

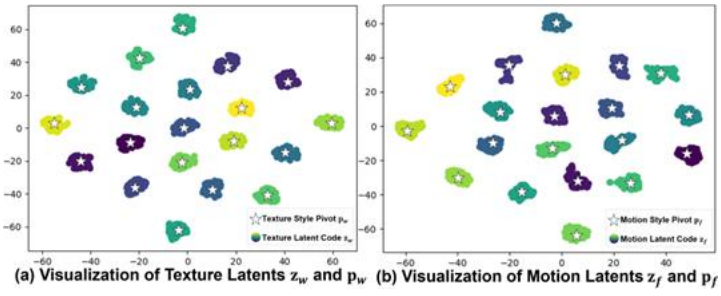


图 10　来自 20 个个体的隐特征空间 t-SNE 可视化

Method	LVE↓ 10^{-2}mm	MVE↓ 10^{-2}mm	FDD↓ 10^{-3}mm
FaceFormer	1.80	2.94	1.68
CodeTalker	1.83	2.80	1.38
FaceDiffuser	1.53	2.38	1.64s
TexTalker	**1.49**	**2.34**	**1.20**

表 2　几何精度比较

基于高效的低维隐式表示，我们训练一个基于Transformer的隐扩散模型$\mathcal{F}$学习语音驱动几何纹理。具体来说，z_f和z_w首先被拼接在一起构成一个动画样本$X^0 = [z_f, z_w]$。随后，向干净样本X^0中逐步加入高斯噪声得到$X^n (n = 1,2,\dots,N)$。$\mathcal{F}$学习在音频特征的引导下从噪声样本中降噪生成干净样本。我们使用预训练的HuBERT编码器生成音频特征。为了实现更好的时序稳定性，同时建模长距离关联关系，模型同时生成一段时序窗口内的特征：

$$\hat{X}^0_{-T_p:T_w} = \mathcal{F}\left(X^n_{0:T_w}, X^0_{-T_p:0}, A_{-T_p:T_w}, n\right).$$

其中我们不仅输入噪声样本$X^n_{0:T_w}$，还输入长度为T_p的来自上个窗口的干净样本$X^0_{-T_p:0}$和降噪时间步n。我们使用样本重建损失训练模型：

$$\mathcal{L}_{\mathcal{F}} = \left|\left|\hat{X}^0_{-T_p:T_w} - X^0_{-T_p:T_w}\right|\right|_2.$$

4.3 解耦的说话与褶皱风格注入

为了实现风格控制，我们的核心思想在于学习到的离散码本存储了动画原语，同时有相似风格的隐特征共享相似的原语特征，因此相同风格的隐特征应当在隐空间中聚类在一起。如图10所示，我们观察到来自相同受试者的面部运动特征与纹理变化特征都被聚类在一起。因此我们将聚类中心作为风格特征：$p = 1/T\sum_{t=1}^{T} z_t$.

这种设计的额外好处是风格特征和隐特征来自同一个空间，因此我们的动画生成模型$\mathcal{F}$只需要学习生成与身份无关但与音频相关的特征偏移量$\Delta z = z - p$，而风格可以被无缝注入，同时进行灵活的风格替换。在此基础上，$\mathcal{F}$生成偏移样本$X'_0 = [\Delta z_f, \Delta z_w]$表示为：

$$\hat{X}'^0_{-T_p:T_w} = \mathcal{F}\left(X'^n_{0:T_w}, X'^0_{-T_p:0}, A_{-T_p:T_w}, n\right).$$

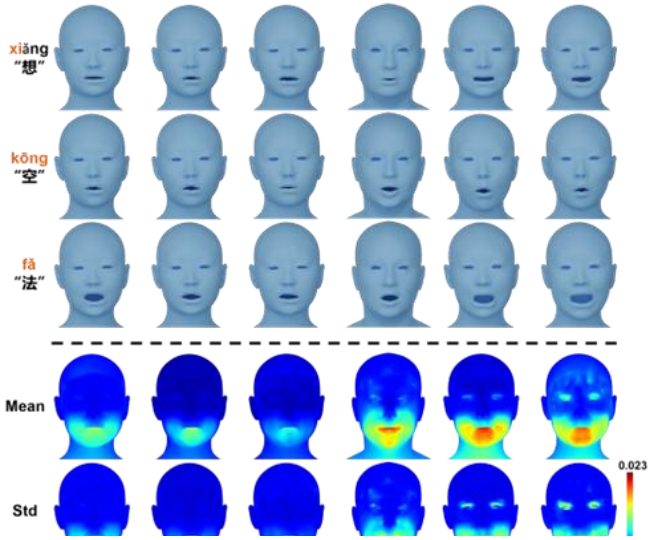


图 11　面部运动生成结果比较。下半部分展示了生成结果在整个序列上的动态统计信息。

最终在推理阶段，可以使用风格i的说话风格特征$p_{f,i}$来构建面部运动特征: $\hat{z}_f = p_{f,i} + \Delta\hat{z}_f$，用风格j的纹理风格特征$p_{w,j}$来构建面部纹理特征: $\hat{z}_w = p_{w,j} + \Delta\hat{z}_w$。利用训练好的生成器$\mathcal{G}_w$和$\mathcal{G}_f$可以从隐式特征重建出具有风格i的面部运动贴图序列和具有风格j的褶皱贴图序列，并从中恢复个性化的动态人脸模型和动态纹理贴图。

4.4 相关实验

数据集　我们将TexTalk4D分为来自85个受试者的80分钟训练集TexTalk4D-Train、来自5个受试者的5分钟验证集TexTalk4D-Valid、来自10个训练集中受试者的5分钟测试集TexTalk4D-Test-A、来自10个未知受试者的5分钟测试集TexTalk4D-Test-B。Test-A用于计算客观指标并衡量几何精度，Test-B用于定性分析。

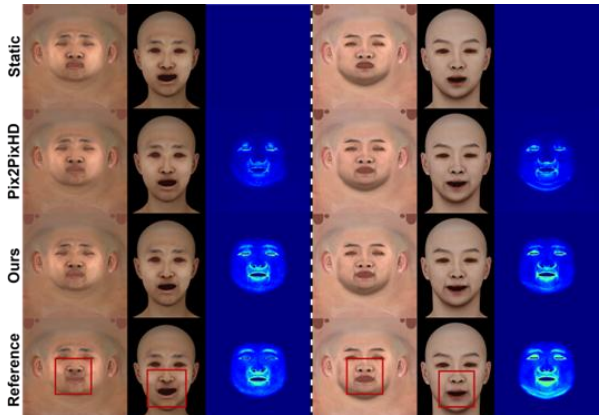


图 12　纹理贴图生成结果比较。第三列和第六列用热图展示了统计热图，颜色越亮代表动态范围越大。

几何质量评估　我们对比了FaceFormer[10]、CodeTalker[11]、FaceDiffuser[12]和DiffPoseTalk[13]。由于DiffPoseTalk只与Flame系数兼容，只进行定性对比。我们计算嘴唇顶点误差（LVE）、全脸顶点误差（MVE）、上脸动态变化（FDD）作为精度衡量指标。比较结果如表2所示，我们的方法优于其他方法。视觉比较结果如图11所示，我们的方法更加匹配真实嘴型。同时能够生成更大动态范围的结果，包括难以生成的自然的眨眼动作。

纹理质量评估　参考Zhao等人[14]，我们训练一个Pix2PixHD学习从面部运动贴图生成纹理贴图，如图12所示。表3展示了定量对比结果，我们方法显著优于静态纹理和Zhao等人的方法。

几何-纹理一致性评估　我们设计了一个user study来进行一致性比较。志愿者从真实感和一致性的角度评估生成的人脸动画的质量，从1-5进行打分。如表3所示，我们的方法生成的动画更加真实，一致性更强。

Meth.	Test-A			Test-B	
	PSNR	SSIM	LPIPS	Real.	Con.
Static	39.79	0.967	0.0146	3.10	2.65
P2PHD	42.34	0.981	0.0187	3.91	3.78
Ours	**44.13**	**0.985**	**0.0101**	**4.13**	**3.97**

表 3　纹理比较。我们在 TexTalk4D-Test-A 上计算指标，在 TexTalk4D-Test-B 上进行 User Study。

五、总结

本文提出了首个基于三维高斯的动态人脸重建方法Topo4D，能够高效重建扫描级几何资产和8K分辨率毛孔级细节纹理贴图。基于Topo4D，我们提出了一个全新的高多样性扫描级精度语音-网格-纹理对齐数据集TexTalk4D，包含100个受试者的说话数据，尤其提供8K动态贴图，能够表现动态面部褶皱变化。基于此数据集，我们提出了首个语音同时驱动几何和纹理的框架TexTalker，能够实现高一致性的说话人脸动画生成，同时能够解耦地控制说话风格和纹理变化风格。大量充分的实验证明了所提出的方法的有效性。该成果已发表于计算机视觉国际顶级会议ECCV 2024和CVPR 2025。

责任编委　张　青

参考文献

[1] Li X, Cheng Y, Ren X, et al. Topo4D: Topology-Preserving Gaussian Splatting for High-fidelity 4D Head Capture[C]//ECCV, 2025: 128-145.

[2] Luiten J, Kopanas G, Leibe B, et al. Dynamic 3d gaussians: Tracking by persistent dynamic view synthesis[J]. arXiv preprint arXiv:2308.09713, 2023.

[3] Feng Y, Feng H, Black M J, et al. Learning an animatable detailed 3D face model from in-the-wild images[J].TOG, 2021.

[4] Lei B, Ren J, Feng M, et al. A hierarchical representation network for accurate and detailed face reconstruction from in-the-wild images[C]//CVPR. 2023: 394-403.\

[5] Xiao Y, Zhu H, Yang H, et al. Detailed facial geometry recovery from multi-view images by learning an implicit function[C]//AAAI. 2022.

[6] Bai Z, Cui Z, Rahim J A, et al. Deep facial non-rigid multi-view stereo[C]//CVPR. 2020: 5850-5860.

[7] Slossberg R, Jubran I, Kimmel R. Unsupervised high-fidelity facial texture generation and reconstruction[C]//ECCV, 2022.

[8] Zhang L, Zeng C, Zhang Q, et al. Video-driven neural physically-based facial asset for production[J].TOG, 2022.

[9] Esser P, Rombach R, Ommer B. Taming transformers for high-resolution image synthesis[C]//CVPR. 2021: 12873-12883.

[10] Fan Y, Lin Z, Saito J, et al. Faceformer: Speech-driven 3d facial animation with transformers[C]//CVPR. 2022: 18770-18780.

[11] Xing J, Xia M, Zhang Y, et al. Codetalker: Speech-driven 3d facial animation with discrete motion prior[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 12780-12790.

[12] Stan S, Haque K I, Yumak Z. Facediffuser: Speech-driven 3d facial animation synthesis using diffusion[C]//Proceedings of the 16th ACM SIGGRAPH Conference on Motion, Interaction and Games. 2023: 1-11.

[13] Sun Z, Lv T, Ye S, et al. Diffposetalk: Speech-driven stylistic 3d facial animation and head pose generation via diffusion models[J]. TOG, 2024, 43(4): 1-9.

[14] Li J, Kuang Z, Zhao Y, et al. Dynamic facial asset and rig generation from a single scan[J]. TOG, 2020, 39(6): 215:1-215:18.

李炫辰

上海交通大学人工智能研究院博士研究生，导师为晏轶超副教授，主要研究方向为三维数字人脸重建、驱动、生成。

Email:lixc6486@sjtu.edu.cn

程宇豪

上海交通大学人工智能研究院博士研究生，导师为戴琼海院士和晏轶超副教授，主要研究方向为隐式数字人脸的生成/编辑以及基于网格的三维数字人脸的重建/驱动。

Email: chengyuhao@sjtu.edu.cn

晏轶超

上海交通大学人工智能研究院副教授，博士生导师。主要研究方向为 AIGC 及三维数字人技术，发表包括 TPAMI、CVPR、NeurIPS 在内的论文 40 余篇。先后主持国家自然科学基金青年项目、CCF-阿里巴巴青年科学家基金等项目 8 项。曾入选上海市海外高层次人才计划，获 2020 年度中国图像图形学学会优秀博士论文奖。

Email:yanyichao@sjtu.edu.cn

苏州科技大学胡伏原教授访谈

2024 年 12 月 12 日，《CCF-CV 专委简报》在线采访了苏州科技大学博士生导师 胡伏原教授。下面是采访实录。

问题 1：胡老师，您好！首先，请您分享一下您的个人学习和研究经历。

大家好！我本科毕业于长安大学，硕士和博士师从西北工业大学张艳宁教授，2007 年博士毕业。在 2006 年到香港城市大学从事研究助理工作，2007 年到 2008 年在比利时 Vrije 大学从事博士后研究工作，2014 年到 Adelaide 大学高访。2009 年回国，一直在苏州科技大学从事教学和科研工作。

问题 2：您的研究工作深耕于多维图像处理、连续学习及应用研究工作。请问，能介绍下您及团队在这方面的成果么？

多源图像处理的研究最早源于 2003 年多源图像处理，当时将可见光和红外视频进行目标检测、识别和跟踪融合，希望可以提升不同天气条件下，目标检测跟踪的精度，该项目成果获得陕西省科学技术奖。随后，从事多相机融合及在图像中的测量研究，在 2010 年发布了苏州市古城街景数据（国内腾讯、百度、高德等都没有发布相应产品），并成为了苏州基础地理数据库重要组成部分。随后在街景数据上进行量测、可编辑等应用研究，开启了广告管理、规划等行政审批足不出户的新模式，相关成果获得江苏省科学技术奖。

随着人工智能技术发展，视频图像、语音和文本应用也越来越广泛。结合长三角智能制造的智改数转需求，人工智能技术应用需求越来越急迫。在现实世界中，数据分布是动态变化的，传统的机器学习模型通常无法适应这种变化，导致模型性能下降。为了提升模型的适应性和鲁棒性，开始从事连续学习及应用研究。目前主要是从事域变化连续学习、小样本连续学习和多模态数据连续学习等，前期相关成果已经在 CVPR、AAAI、TMM 等期刊和会议上发表，成果也汇编为专著《Continual Artificial Intelligence towards Changing Environment》，并获得国家科学技术学术著作出版基金项目资助出版和中国科技产业化推进会科技创新奖一等奖。项目成果现已经应用到开放环境中的动态目标持续检测和识别中，后期期待在机器人，特别是人形机器人中进行推广应用。

问题 3：在您的众多研究成果之中，请问哪一项是您个人最引以为傲的成就？

引以为傲的成就谈不上，我的大部分成果主要是应用需求牵引，也正是我着力解决实际问题、从实际中寻找科学问题，让我有幸参与了苏州多个过亿信息化项目的研究和实施。我感觉有两个成果形成过程记忆非常深刻。一个是硕博连读期间的多源图像融合项目，这个工作不仅仅是在张艳宁导师指引下学习了新的知识，更重要的是将研究的理论、模型应用到钟楼、学校以及无人机航拍视频等众多实际场所中，提升了我的独立科研能力和项目实战管理能力。项目实施过程中，既有白天+黑夜 24 小时不间断的实验室验证实验，也有冰寒雪冻

和春暖花开的野外实验，这是我第一次带领 10 个人的团队做理论和应用相结合的研究，获得知识的快乐和弱小目标跟踪成功的喜悦，深深刻入我的脑海，记忆犹新。另外一件事，是我 2009 年刚刚到苏州，接的电池正负极和焊点异常检测的项目。该项目现在来说难度不大，但是当时刚刚工作，从零到有完成包括流水线装置设计制造、算法设计调试等，每天睡眠时间不超过 4 小时，不到一个月完成了一整套设备的研制和上线。产品应用后每条产线至少节约了 4 个工人，该产品现在已经服役超过 15 年了。这两个项目的实施对我科研之路奠定了良好的基础，也让我深刻体会到如何从实战中发现科学问题，我后续国家基金项目选题都是来自实际需求。

问题 4：您在成果转化方面也有很多的成果，您主持完成横向项目 20 余项，获得中国科技产业化促进会科技创新奖一等奖、省部级科技进步奖二等奖 3 项，三等奖 4 项，请问您能介绍下在成果转化方面的经验么？

经验谈不上。作为农村孩子，从小玩到大，喜欢做做实际的事情，脚踏实地地解决实际问题。同时，我所在的学校平台不高，硕士研究生数量也不多。但是，我发现为数不多的硕士生，只要和他们经常讨论问题，他们冲劲很足，实际上也很能干，也很愿意吃苦。曾记得，2013 年左右，体育局一个小项目，通过 APP 估计市民在环城河步道的步数发放小礼品（当时计步软件应用还很少），来调动苏州市民运动积极性。研究生和我年轻时一样，为了计步准确，夜以继日的在环城河旁边测试，不到一个月把整个环城河步道中涉及到的过桥洞信号弱、骑行跑步、静止甩手机等计步不准确问题都一一解决了，后期市民都觉得计步很准确，无一人投诉。正是有这帮年轻学生的支持，有苏州智造之城的土壤，让我们团队更有动力将研究成果进行转化，为地区经济发展做一些力所能及的事情。

问题 5：您在教学方面成果颇丰，曾获得省教学成果二等奖 1 项，是国家一流专业负责人，江苏省“333 高层次人才培养工程”中青年科技带头人、江苏高校“青蓝工程”教学团队负责人、中青年学术带头人，并入选江苏省“六大人才高峰”人才培养对象。您能介绍下在教学方面的成果及您的教学理念么？

教学是个良心活，做得好受益学生很多，桃李满天下就是我们最大的收获和快乐。也正是因为有希望学生能够快速成长的育人初心，我常常把自己做的项目实例在课堂上讲，也经常出一些实际的课题，启迪学生思维。我的教学一直与实际应用结合，将实际需求导入课堂，希望学生能够从工程真境中寻找所需的科技前沿知识。通过“产教融合导入需求，科技融汇碰撞灵感”的理念，启迪学生思维、淬炼学生能力，让学生在实境中学习实践，实实在在地体会到学有所用。2017 年度江苏省教学成果奖和 2024 年度中国图象图形学会教育教学成果奖都是通过践行产教融合、科教融汇的理念获得的。

问题 6：您认为科研成果和教学成果是如何相辅相成的？在这方面，您对其他青年科研工作者有什么建议呢？

将科研成果，包括工程项目和前沿项目成果引入教学中，能丰富教学内容，使教学更具深度和广度，激发学生兴趣。教学过程中需要我们对知识，特别是工程项目进行梳理凝练和对前沿成果进行总结梳理，更有逻辑地解释复杂概念和回答学生问题，这些都能够更好地获得新的研究灵感，提升自身科研能力。

谈不上建议，我个人的体会，大家一定要合理安排时间，做脚踏实地的科研，解决实际问题的科研，从实际需求中发现科学问题；同时，要将最新研究成果引入课堂，提升教学质量，也可以吸纳到更多优秀本科生到团队做科研和实践。用严谨求实的态度，精益求精的作风，创新实践的方法做教学和科研，一定可以做到两者的有机结合，实现教育质量和科研水平的双重提升。

问题 7：您在构建科研团队方面有什么经验？在青年教师与研究生的培养与管理上，您采取了哪些独到而高效的策略？能否分享一些您在实践中总结出的优秀做法，以资同行借鉴与学习？

科研团队的构建，很多老师比我有经验，就不班门

弄斧了。青年老师我觉得还是要挖掘他们的潜力，每个老师偏好不一样，不是所有博士生毕业到高校，都能够很快独立科研。每位老师的成长都希望被认可，都希望在学院或者团队有一个属于自己的位置。同时，每个老师都有缺点或者做错事情的时候，要包容。我一般是从项目和合作带学生中观察年轻老师的能力以及想法，每个老师有侧重，有的偏重教学研究、有的偏重理论研究，还有的侧重成果转化和实际项目落地。发挥每个老师长处，让他们都能在工作中有较大收获；同时也考虑他们未来成长和个人收益。对于研究生，根据学生兴趣分为理论研究和工程项目实施两类来指导，同时也确保学生生活费，让他们基本生活无忧，开心快乐做科研。我这边学生读博士，基本上都是我亲自联系导师，带着他们到对应高校或者研究所博导那边交流，成功率很高，也为他们骄傲。他们很多读博士的时候，也反馈我们团队，经常协助指导研究生。

问题 8：在繁忙的工作之余，您有哪些爱好，以给自己放松和充电呢？同时，您又是如何平衡工作与个人家庭生活，确保两者和谐共生的？

在工作之余，我喜欢通过阅读和运动来放松和充电。阅读让我明事理，拓宽视野也为我提供了新的灵感和动力；运动则帮助我保持身体健康，释放压力。健康的身体是一切事情的保障，大家一定要多抽时间锻炼身体。

在平衡工作与家庭生活方面，我始终坚持“时间管理”和“优先级排序”的原则。尽量高效完成任务，留出足够的时间陪伴家人。家庭是我工作最重要的保障，我会常与家人沟通，相互尊重，理解彼此，确保工作与家庭生活的和谐共生，实现两者的共同发展。

问题 9：如果吐露研究工作者的心声，您最想说的是什么？

作为科研工作者，要不忘初心，一直保持对科学的热爱、对真理的追求和对未来的期待。尽管经常遇到失败，但我们依然要继续奋斗、要乐于奋斗。唯有坚持和努力，我们才能完成使命，达成梦想。

责任编委 余烨 赵振兵

胡伏原

教授/博士，博士生导师，国家一流专业负责人。现任苏州科技大学电子与信息工程学院院长。江苏省“333 高层次人才培养工程”中青年科技带头人、江苏高校“青蓝工程”教学团队负责人、中青年学术带头人，并入选江苏省“六大人才高峰”人才培养对象。现为中国体视学学会常务理事，CCF 计算机视觉专委会委员，中国图象图形学会成像探测与感知专委会委员。

一直以来从事多维图像处理、连续学习及应用研究工作，已主持国家自然科学基金面上项目、省重点研发计划等科研项目 10 多项，主持完成横向项目 20 余项。获得省教学成果二等奖 1 项，中国科技产业化促进会科技创新奖一等奖、省部级科技进步奖二等奖 3 项，三等奖 4 项，在《IEEE Trans. on PAMI》、《IEEE Trans. on Multimedia》、AAAI、CVPR 等重要国内外学术期刊发表学术论文 100 余篇，授权发明专利 26 件。

委员好消息

✪ 2024 年 12 月 17 日，中国自动化学会发布了 2024 中国自动化学会科学技术奖评审结果公告，共授奖 124 项，CCF-CV 专委会 9 位执行委员的项目获奖：北京大学**林宙辰**等完成的“非完备情况下的表示学习理论与方法” 和北京空间飞行器总体设计部**王大轶**等完成的“空间无人系统资源强受限下的可观测性理论及方法”获自然科学一等奖，中国科学院信息工程研究所**张晓宇**等完成的“网络空间威胁模式分析与识别”、大连理工大学**杨鑫**等完成的“复杂视觉条件下的机器人环境感知与目标识别”、厦门大学**曲延云**等完成的“高阶结构正则的多视图机器学习理论与方法”获自然科学二等奖，中国科学院自动化研究所**王坤峰**等完成的“复杂交通环境自动驾驶可信感知理论与方法”、中国科学院计算技术研究所**张杰**等完成的“城市道路震后可穿越性规划与最优协同控制研究”获自然科学三等奖，航天宏图信息技术股份有限公司**王涛**等完成的“复杂城市场景高精高效三维感知关键技术及应用”获科技进步一等奖，山东大学**张伟**等完成的“数智化高效喷涂作业关键技术及应用”获科技进步二等奖。

✪ 2025 年 1 月 3 日，中国计算机学会公布了 2024 年度 CCF 夏培肃奖获奖名单，共 2 人上榜，CCF-CV 专委会执行委员、西北工业大学**张艳宁**入选。张艳宁教授是计算机视觉领域的杰出学者，在天基空间环境光学探测方面做出了突出贡献，彰显了女性科技工作者的学术领导力和影响力。

✪ 2025 年 1 月 8 日，中国自动化学会发布了 2024 中国自动化学会研究生论文工程评价结果公告，CCF-CV 专委会 2 位执行委员指导的研究生学位论文入选：北京大学**林宙辰**指导的《隐式均衡模型的设计和加速》入选博士论文一等学位论文，中国石油大学（华东）**刘伟峰**指导的《面向小样本分类问题的任务选择方法研究》入选硕士论文。

✪ 2025 年 2 月 20 日，中国自动化学会发布了 2024 中国自动化学会高等教育（本科、研究生）教学成果评价结果公告，CCF-CV 专委会 2 位执行委员的成果入选：北京科技大学**樊彬**等完成的“思政引领，一本双驱 自动化类(智能方向)创新人才培养体系研究与实践”获一等贡献教学成果，四川大学**胡鹏**等完成的“发展新质生产力视阈下 ‘产-教-研’ 共同体育人模式的创新与实践”获三等贡献教学成果。

✪ 2025 年 2 月 20 日，中国人工智能学会公布了 2024 年度 CAAI 激励计划入选名单，本年度共有 31 个项目入选，CCF-CV 专委会 2 位执行委员获奖：西安交通大学**白慧慧**主持的“ ‘AI+交通’ 分阶段课赛协同的创新人才培养体系建设”和 CCF-CV 专委会执行委员、西安电子科技大学**董伟生**参与的“人工智能交叉领域创新人才培养新模式探索与实践”入选 CAAI 教学成果激励计划一类成果。

✪ 2025 年 2 月 20 日，中国人工智能学会发布了 2024 年度吴文俊人工智能科学技术奖奖励公告，决定授 53 项 2024 年度“吴文俊人工智能科学技术奖”，CCF-CV 专委会 7 位执行委员完成的 5 个项目入选：北京航空航天大学**黄迪**、**王蕴红**和中国科学院自动化研究所**雷震**等完成的“复杂视觉任务的高效表征学习”获自然科学一等奖，京东科技信息技术有限公司**刘武**等完成的“多模态交互式数字人关键技术及产业应用”获科技进步特等奖，天津大学**朱鹏飞**等完成的“低空智能感知关键技术与应用”、浙江大学**赵洲**等完成的“智慧司法智能化支撑平台与示范应用”获科技进步一等奖，北京邮电大学**马**

占宇等完成的“可信多模态数据流通关键技术及产业化应用”获科技进步二等奖。

✪ 2025 年 3 月 7 日，中国电子学会发布 2024 中国电子学会科学技术奖励公告，共授奖 139 项，CCF-CV 专委会 9 位执行委员获奖：中山大学**操晓春**、中国科学院信息工程研究所**任文琦**等完成的“低质视觉数据的特征发掘与约束寻优”、清华大学**鲁继文**、**段岳圻**等完成的“高效稳健的视觉信息分析与识别理论方法”获自然科学一等奖，北京理工大学**杨健**等完成的“面向胸腹部手术导航的图像高分辨重建融合关键技术及系统”、中国科学院自动化研究所**张兆翔**等完成的“大规模复杂科技大数据智能分析与服务技术及应用”获科技进步一等奖，北京理工大学**付莹**、北京交通大学**白慧慧**、北京大学**施柏鑫**获青年科学家奖。

责任编委 刘海波

基于神经辐射场的渲染方法及开源代码

香港城市大学 付陈平 大连理工大学 樊鑫

神经辐射场（Neural Radiance Fields, NeRF）及其变体为渲染社区带来全新的应用体验，其发展潜力得到研究人员的热烈关注。简单而言，NeRF 将场景表示为由多层感知器存储的连续辐射场，通过积分密度和辐射渲染新视图。近几年，研究人员在多样渲染场景中对 NeRF 技术展开广泛探索。从无界场景渲染，到运动物体场景渲染，再到野外场景渲染，NeRF 均取得令人欣喜的场景重建效果。鉴于此，本文立足 NeRF 技术，介绍几种渲染方法，同时提供相应代码链接，供读者参考。

1、MS-NeRF 方法

介绍：目前，NeRF 技术在各种运动、低光等场景中展现出优越的渲染性能。然而，该技术尚无法在带有镜子的场景中保持优良的渲染效果。渲染过程中，NeRF 会主动假设目标场景存在多视图一致性。然而，当空间中有镜子，且允许视点围绕场景 360 度移动时，由于镜子表面和它反射的虚拟图像只有在很小的视野范围内可见，因此镜子的前后视图之间存在不一致性。此时，需要手动标记反射表面，以避免陷入次优收敛。

针对上述问题，本文提出一种基于 NeRF 的多空间方法--MS-NeRF。MS-NeRF 在场景 360 度高保真渲染中自动处理镜面及其类镜面物体，无需任何人工标记。如图 1 所示，MS-NeRF 将场景空间视为多个虚拟子空间的组合，不将其视为一个单独的空间，这些虚拟子空间的组成随着位置和视点方向的变化而变化。此种多空间分解方法可以成功处理复杂的反射和折射场景。本文通过设计一个低成本多空间模块，并用该模块替换原来 NeRF 方法输出层，实现空间分解。此外，该多空间模块是一个通用模块，可与现有 NeRF 方法结合，提高它们的镜面渲染效果。

论文地址：
https://ieeexplore.ieee.org/document/10878467

代码地址： https://zx-yin.github.io/msnerf/

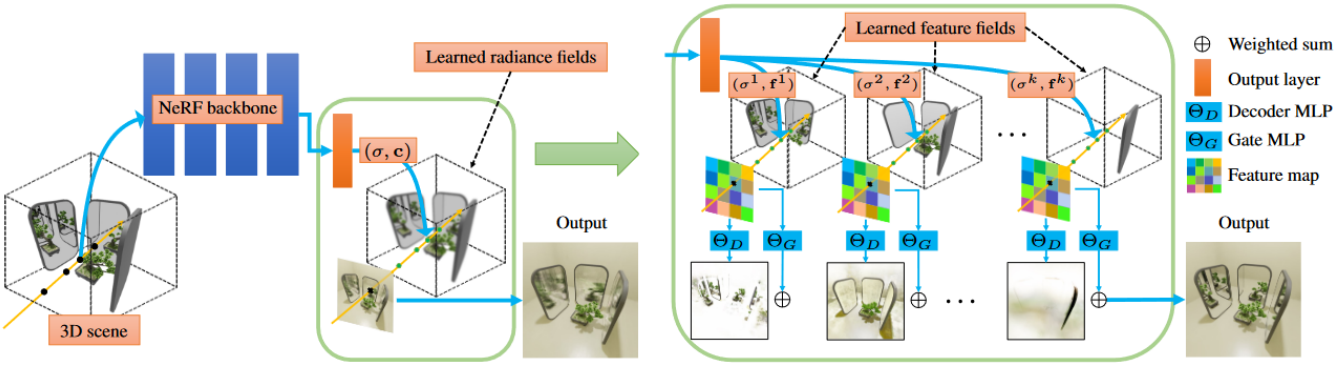


图 1 MS-NeRF 渲染方法框架图

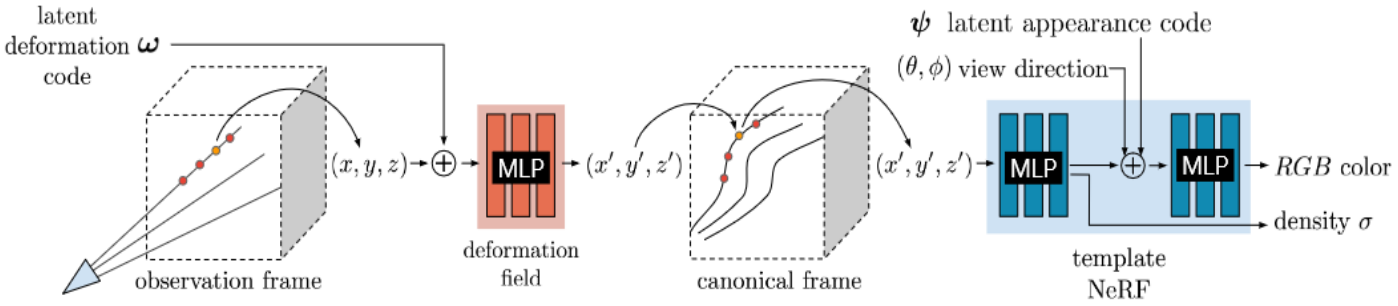


图 2 Nerfies 渲染方法框架图

2、Nerfies 方法

介绍：目前，高质量 3D 人体扫描技术已经取得很大进展，但需要一个具有许多同步灯光和摄像机的专门实验室才可以保证 3D 人体扫描的正常运行。相较于上述方式，研究人员更希望只需挥动手机、相机就能捕捉他人或自己的 3D 建模，这种方式将极大减轻 3D 建模技术的访问和应用。

用手持相机对人进行 3D 建模特别具有挑战性，因为：（1）非刚性——待建模人无法保持完美静止，（2）待建模人往往带有各种违背渲染假设视图一致性的材料，如头发、眼镜和耳环。本文提出一种解决这两个挑战的方法—Nerfies。Nerfie 通过泛化神经辐射场对形状变形进行建模。该技术可以从短视频中恢复高保真 3D 重建，提供自由视角的可视化，同时可以准确捕捉头发、眼镜等其他复杂与视角相关的材料。如图 2 所示，该方法通过优化额外的连续体积变形场来增强神经辐射场，该场将每个观测点扭曲为标准的 5D NeRF。这些类似 NeRF 的变形场容易出现局部最小值，因此，本文提出一种基于坐标模型的由粗到细的优化方法，实现鲁棒优化。此外，本文通过将几何处理和物理模拟的原则适应于类似 Nerf 的模型，提出变形场的弹性正则化，进一步提高 Nerfies 鲁棒性。

所提方法可以将随意拍摄的自拍照片/视频转变为可变形的 NeRF 模型，允许从任意视角对主题进行逼真的场景渲染。本文利用两部手机收集时间同步数据构建在不同视角下具有相同姿态的训练/验证图像，用于评估所提方法。Nerfies 成功重建非刚性变形场景，并以高保真度再现未出现过的视图。

论文地址： https://arxiv.org/pdf/2011.12948

代码地址： https://nerfies.github.io/

责任编委 李策 王田

付陈平

博士后，香港城市大学计算机科学学院，研究方向为计算机视觉，目标检测，图像增强。

樊 鑫

博士生导师，大连理工大学国际信息与软件学院从事教学与科研工作，担任软件学院院长。研究方向为计算机视觉与图像处理、医学影像分析。
个人主页：http://faculty.dlut.edu.cn/Xin_Fan/zh_CN/index.htm

人体动作生成数据集

北京航空航天大学 曹哲骁 王田

人体动作生成（Human Motion Generation）旨在通过算法自动生成逼真、多样化且符合语义的人体动作序列。其核心是通过条件信号（如文本、音频、场景或图像）控制生成过程，使动作与输入条件高度一致。任务的主要挑战包括：需满足生物力学约束（如关节角度、物理平衡），生成动作自然；动作与文本/音频的语义一致性，符合多模态对齐；在长序列生成时保持时序连贯性，避免动作断裂；多人或人与物体交互的动态协调。

人体动作生成正从单一模态向多模态、单人向多人交互演进，未来或成为虚拟数字人技术的核心引擎。本文重点介绍了人体动作生成领域一些常见的公开数据集。

1. KIT

KIT 由德国卡尔斯鲁厄理工学院（KIT）的 Matthias Plappert、Christian Mandery 和 Tamim Asfour 团队于 2016 年发布。核心目标是建立人体运动数据与自然语言描述的关联，推动动作语义理解、机器人动作生成等领域的研究。数据集包含 3,966 段运动数据和 6,278 个自然语言标注。

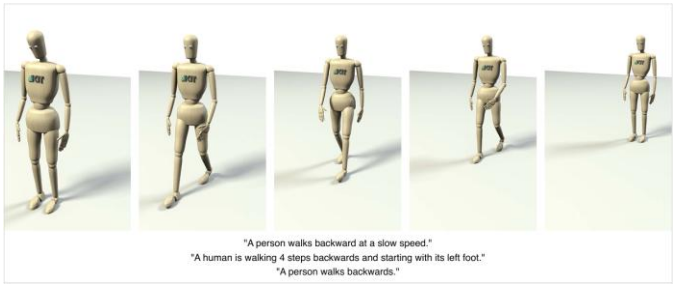


图 1 KIT 数据集中动作与描述示例

数据集下载地址

https://motion-annotation.humanoids.kit.edu/dataset/

相关论文链接：

The KIT Motion-Language Dataset

https://matthiasplappert.com/publications/2016_Plappert_Big-Data.pdf

2. HumanML3D

HumanML3D 由阿尔伯塔大学 Chuan Guo 团队发表于 CVPR 2022。数据集整合了 HumanAct12 和 AMASS 数据集，并扩展了 KIT-ML 数据集，是一个大规模、多样化的 3D 人体运动-语言多模态数据集，旨在通过自然语言描述与 3D 人体动作的精确关联，支持文本驱动动作生成、动作检索等任务。数据集包含 14,616 段运动数据和 44,970 个自然语言标注。

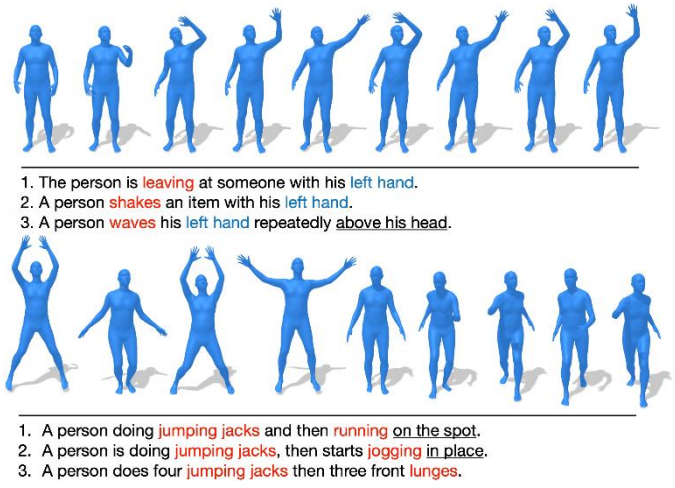


图 2 HumanML3D 数据集中动作与描述示例

数据集下载地址

https://github.com/EricGuo5513/HumanML3D

相关论文链接：

Generating Diverse and Natural 3D Human Motions From Text

https://openaccess.thecvf.com/content/CVPR2022/papers/Guo_Generating_Diverse_and_Natural_3D_Human_Motions_From_Text_CVPR_2022_paper.pdf

3. BABEL

BABEL (Action-Antification Benchmark for Evaluating Localization)由马克斯·普朗克智能系统研究所（MPI）的 Abhinanda Punnakkal 等人发表于 CVPR 2021。基于 AMASS 数据集整合了 60 余个运动捕捉数据库，通过人工标注扩展动作语义标签。旨在为 3D 人体动作提供细粒度的时间动作定位和多标签分类基准，支持动作识别、时序分割、跨模态对齐等任务。数据集包含 13,220 段动作数据，28,055 个段落标签和 63,353 个帧级标签。

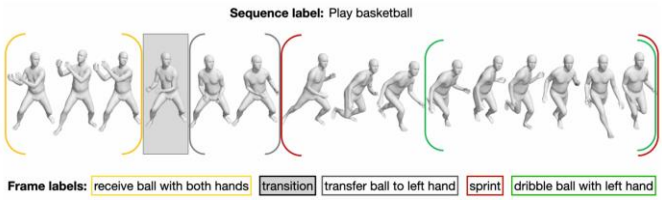


图 3 BABEL 数据集中动作与描述示例

数据集下载地址

https://babel.is.tue.mpg.de/

相关论文链接：

BABEL: Bodies, Action and Behavior with English Labels

https://arxiv.org/pdf/2106.09696

4. AIST++

AIST++由 Google Research、加州大学伯克利分校等团队联合开发，相关论文发表于 ICCV 2021。旨在提供音乐驱动的 3D 舞蹈动作生成与多视角人体姿态分析的基准数据集，支持舞蹈动作合成、跨模态（音乐-动作）关联研究。基于原始 AIST 舞蹈视频数据库构建，通过多视角重建技术生成 3D 标注。数据集包含 1,408 段动作数据和 60 首不同节奏的独立音乐。

数据集下载地址

https://google.github.io/aistplusplus_dataset/

相关论文链接：

AI Choreographer: Music Conditioned 3D Dance Generation with AIST++

https://google.github.io/aichoreographer/

责任编委 樊鑫 贾同

曹哲骁

北京航空航天大学人工智能学院，博士研究生，研究方向为自监督学习、计算机视觉等。

王 田

副教授，博士生导师，北京航空航天大学人工智能学院，院长助理，研究方向为模式识别与智能系统，图像异常检测。

好 文 推 荐

厦门大学“CamoDiffusion: Camouflaged Object Detection via Conditional Diffusion Models”的最新成果发表在 IEEE TPAMI 2025。

论文：Zhongxi Chen, Ke Sun, Xianming Lin, Rongrong Ji. Conditional Diffusion Models for Camouflaged and Salient Object Detection, IEEE TPAMI, 47(4), 2833-2848, 2025.

在伪装目标检测任务中，由于伪装目标与其环境极为相似，识别它们仍然是一个重大挑战。现有分割方法难以准确区分目标与背景，并且依赖像素级概率计算，导致过度自信的错误预测，影响检测准确性。

为了解决上述问题，本文将伪装目标检测(Camouflaged Object Detection，COD) 视为基于扩散模型的条件掩膜生成任务。基于此范式提出了一个名为 CamoDiffusion 的新框架，该框架利用扩散模型的去噪过程，以迭代方式逐渐消除初始噪声与真实掩码之间的偏差，并将输入图像作为辅助条件进行引导。具体而言，如图 1 所示，本文通过利用图像先验对每个后续步骤进行条件约束，逐步生成预测结果。框架主要由自适应 Transformer 条件网络(Adaptive Transformer Conditional Network， ATCN) 和基础去噪网络(Denoising Network，DN)组成。本文设计并探讨了两个阶段对模型进行优化：(a) 训练阶段：给定一个输入图像，初始掩码表示目标的真实掩码，对其施加噪声生成噪声掩码，前向过程中根据概率分布生成不同程度的噪声掩码，训练模型学习去噪过程，通过 ATCN 接收图像作为条件信息，提取特征并指导去噪过程，DN 利用 ATCN 提取的条件信息生成去噪后的掩码并与真实掩码计算损失；(b) 采样阶段：ATCN 接收图像并提取条件特征，DN 依据 ATCN 的指导对噪声掩码去噪，这一过程持续迭代，逐步减少噪声，使预测掩码趋近真实目标，去噪完成后得到最终的伪装目标预测掩码。

CamoDiffusion 算法在 CAMO，COD10K，NC4K 数据集上的广泛实验展现出 CamoDiffusion 相较于现有方法的独特优势。特别是在最具挑战性的 COD10K 数据集上，该方法在 MAE 方面达到了 0.019。

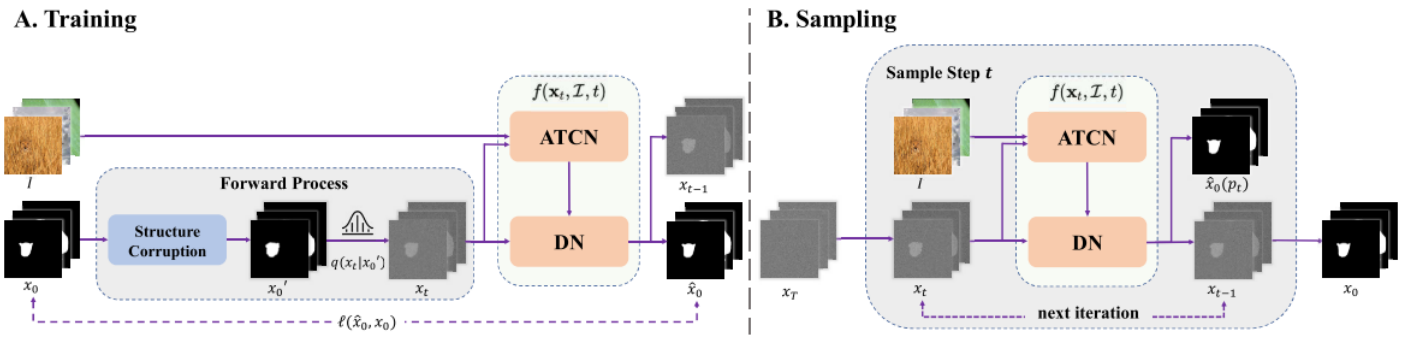


图 1 CamoDiffusion 算法流程图

责任编委 王田 贾同

好 文 推 荐

清华大学的“Text-guided Sparse Voxel Pruning for Efficient 3D Visual Grounding”的最新成果发表在 CVPR 2025。

论文：Wenxuan Guo; Xiuwei Xu; Ziwei Wang; Jianjiang Feng; Chenglu Wen; Jie Zhou. Text-guided Sparse Voxel Pruning for Efficient 3D Visual Grounding, CVPR 2025.

3D 视觉定位（3D Visual Grounding, 3DVG）任务旨在根据自然语言描述在三维场景中定位指定的目标对象。大多方法采用两阶段框架：首先通过 3D 目标检测在场景中找到所有候选物体，然后结合文本信息在第二阶段选出与描述匹配的目标。这种方法虽然直观，但由于两个阶段分别提取特征，存在大量冗余计算，难以满足实际应用中的推理速度要求。为提升效率，单阶段方法直接从点云数据中定位目标物体，将目标检测与语言匹配一步完成。然而，现有单阶段方法大多同样基于点云处理架构，其特征提取需要耗时的最远点采样和近邻搜索等操作。因此当前单阶段方法距离实时推理仍有差距（推理速度不足 6 FPS）。

为了解决上述问题，如图 1 所示，本文提出了一种全新的单阶段 3DVG 框架——TSP3D，即“Text-guided Sparse voxel Pruning for 3DVG”。TSP3D 放弃被现有方法广泛使用的点云处理架构，引入了多层稀疏卷积架构来同时实现高精度和高速推理。三维稀疏卷积架构提供了更高的分辨率和更精细的场景表示。同时，引入基于文本引导的体素剪枝（Text-guided Pruning, TGP）。TGP 的核心思想是赋予模型两方面的能力：(1) 在文本引导下修剪冗余体素来减少特征量；(2) 引导网络将注意力逐渐集中到最终目标上。TSP3D 包含 3 个层次的稀疏卷积和两次特征上采样，因此相应设置了两阶段的 TGP 模块：场景级 TGP (level 3 to 2) 和目标级 TGP (level 2 to 1)。场景级 TGP 旨在区分物体和背景，用来修剪背景上的体素。目标级 TGP 侧重于文本中提到的区域，保留目标对象和参考对象，同时修剪其他区域的体素。根据语言描述逐步修剪对目标定位没有帮助的体素。与之前的单阶段方法相比，TSP3D 实现了最高的推理速度，并且比之前最快的方法快 100% FPS。在 ScanRefer 上 Acc@0.5 领先 +1.13，在 NR3D 和 SR3D 上分别领先 +2.6 和 +3.2。

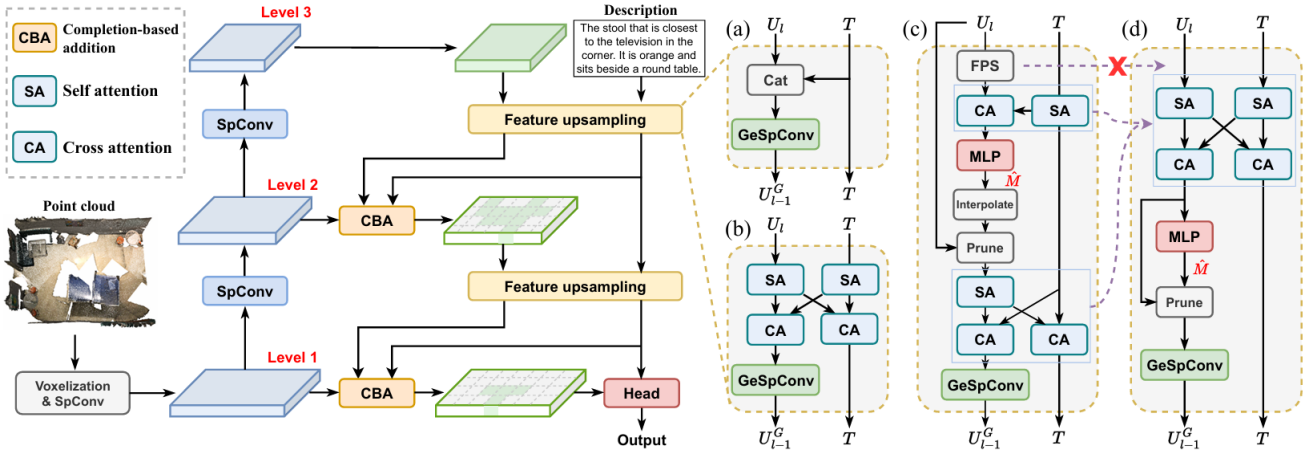


图 1 TSP3D 网络结构流程图

责任编委 贾同 李策

好 文 推 荐

南京航空航天大学脑机智能技术实验室和模式分析与机器智能重点实验室“Tumor Micro-environment Interactions Guided Graph Learning for Survival Analysis of Human Cancers from Whole-slide Pathological Images”的最新成果被 CVPR-2024 收录。

论文：Shao W, Shi Y Y, Zhang D, et al. Tumor micro-environment interactions guided graph learning for survival analysis of human cancers from whole-slide pathological images[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 11694-11703.

在癌症精准医疗领域，患者生存预测是制定个性化治疗方案的核心环节。近年来，基于全切片病理图像（WSI）的深度学习技术为生存分析提供了新思路。然而，现有方法多聚焦于肿瘤区域内的图像块，忽视了肿瘤微环境（TME）中不同成分（如淋巴细胞、间质纤维化）与肿瘤的复杂交互，导致模型预测能力受限。文章通过图学习框架首次将 TME 交互引入生存分析，显著提升了预测精度与可解释性，为癌症预后研究开辟了新方向。

文章提出的图形学习算法（TMEGL）核心突破在于首次系统整合肿瘤微环境的空间交互信息。传统方法通常将 WSI 视为孤立图像块的集合，而 TMEGL 通过以下三方面创新实现了更全面的建模：

文章首先从 WSI 中提取肿瘤、淋巴细胞和间质纤维化三类关键图像块作为图节点，基于空间距离构建拓扑图，直观表征 TME 的空间分布。然后，文章考虑每个节点 TME 结构，提出一种新型图嵌入算法，通过谱聚类量化节点邻域内不同 TME 成分的比例（1-hop 至 3-hop 范围），并设计 KL 散度损失函数，确保嵌入向量保留 TME 的拓扑组织特征。接着采用门控图注意力网络（GGAT），结合门控图卷积（GGC）与图注意力机制（GAT），分别捕捉同类节点间的内部状态差异及跨类节点间的交互（如肿瘤-淋巴细胞抑制关系），动态筛选与生存强相关的交互模式。最后，采用 Cox 比例风险模型，将全局池化后的图特征映射至生存风险评分验证患者分层的显著性。TMEGL 的 C-Index 与 AUC 值均显著优于现有方法（如 WSISA、DeepGraphSurv 等）。

这一工作为计算病理学提供了重要工具，也为探索肿瘤微环境的动态演化打开了新窗口。期待其在临床转化中发挥更大价值，助力癌症精准医疗迈向新高度。

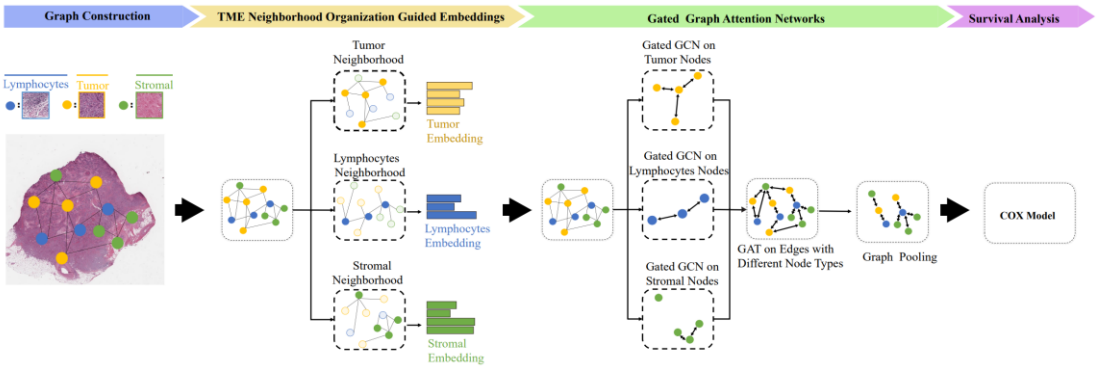


图 1 TMEGL 算法流程图

责任编委 李策 王田

征文通知

1 会议征文

计算机视觉领域相关国内外会议的征文通知如表 1 所示。同时，可继续关注每个会议举办的 workshop 或 special session。

2 期刊征文

计算机视觉领域近期相关期刊专刊的征文通知如表 2 所示，包括 IEEE Journal of Selected Topics in Signal Processing，Pattern Recognition 和 Information Fusion。

3 会议简介

中国模式识别与计算机视觉学术会议 PRCV (Chinese Conference on Pattern Recognition and Computer Vision)，由中国计算机学会（CCF）、中国自动化学会（CAA）、中国图象图形学学会（CSIG）和中国人工智能学会（CAAI）联合主办，定位国内顶级的模式识别和计算机视觉领域学术盛会。

第八届 PRCV 将于 2025 年 10 月 16 日至 10 月 19 日在上海举办，由上海交通大学承办。本届会议将秉持团结模式识别与计算机视觉领域科技工作者的宗旨，进一步推动开放合作，广泛吸引学术界和工业界的人才，提升会议的国际化水平，力求打造一个高品质的学术交流平台。大会的举办将为学术界与工业界提供更多产学研合作机会，推动模式识别与计算机视觉领域的协同创新和可持续发展。

责任编辑：刘帅奇

表 1 计算机视觉领域相关国内外会议

会议名称	会议时间	会议地点	截稿日期	会议网站
ACM MM 2025	2025.10.27-31	Dublin, Ireland	2025.04.04	https://iccv2025.thecvf.com/
NeurIPS 2025	2025.12.02-07	San Diego, United States	2025.05.11	https://neurips.cc/
SiPS 2025	2025.11.01-04	Hong Kong, China	2025.06.01	https://events.polyu.edu.hk/sips2025/home

表 2 计算机视觉领域相关国内外期刊专刊

期刊名称	专刊题目	投稿网址	截稿日期
JSTSP	High-Dimensional Imaging: Emerging Challenges and Advances in Reconstruction and Restoration	https://signalprocessingsociety.org/publications-resources/special-issue-deadlines/ieee-jstsp-special-issue-high-dimensional-imaging-emerging-challenges-and-advances	2025.05.31
PR	Graph Foundation Model for Medical Image Analysis	https://www.sciencedirect.com/special-issue/314872/graph-foundation-model-for-medical-image-analysis	2025.05.01
PR	Beneficial Noise Learning	https://www.sciencedirect.com/special-issue/316469/beneficial-noise-learning	2025.06.15
IF	GenAI for Information Fusion	https://www.sciencedirect.com/special-issue/316104/genai-for-information-fusion	2025.06.30

COMPUTER VISION NEWSLETTER

01 2025
总第 43 期

计算机视觉专委会简报

CCF 计算机视觉专委会