# ICCV 2021 论文整理

[ICCV 2021 论文整理](#iccv-2021-论文整理)  
 [一般的深度学习、机器学习任务](#一般的深度学习机器学习任务)  
 [网络设计](#网络设计)  
 [分类](#分类)  
 [聚类](#聚类)  
 [无监督/自监督](#无监督自监督)  
 [半监督](#半监督)  
 [对抗](#对抗)  
 [网络加速](#网络加速)  
 [模型剪枝](#模型剪枝)  
 [知识蒸馏](#知识蒸馏)  
 [Loss](#loss)  
 [模型量化](#模型量化)  
 [迁移学习](#迁移学习)  
 [可解释性](#可解释性)  
 [持续学习](#持续学习)  
 [主动学习](#主动学习)  
 [元学习](#元学习)  
 [训练](#训练)  
 [深度度量学习](#深度度量学习)  
 [其他](#其他-1)  
 [脉冲神经网络](#脉冲神经网络)  
 [图神经网络](#图神经网络)  
 [数据增广](#数据增广)  
 [NAS](#nas)  
 [视觉表征学习](#视觉表征学习)  
 [DA](#da)  
 [DG](#dg)  
 [生成式](#生成式)  
 [图像生成](#图像生成)  
 [3D生成](#Xa6520535ae51542afb9c91d68d77114e4ff9164)  
 [GAN](#gan)  
 [动作识别](#动作识别)  
 [图像理解](#图像理解)  
 [图像去噪、去光等](#图像去噪去光等)  
 [图像检索](#图像检索)  
 [图像变化理解](#图像变化理解)  
 [图像超分辨率](#图像超分辨率)  
 [图像绘制](#图像绘制)  
 [图像恢复](#图像恢复)  
 [图像重建](#图像重建)  
 [图像质量评估](#图像质量评估)  
 [物体定位](#物体定位-1)  
 [密集预测](#密集预测)  
 [图像转换](#图像转换)  
 [图像匹配](#图像匹配)  
 [图像压缩](#图像压缩)  
 [成像](#成像)  
 [图像全景拼接](#图像全景拼接)  
 [相机定位、姿态回归](#相机定位姿态回归)  
 [图像异常检测](#图像异常检测)  
 [视觉文本识别](#视觉文本识别)  
 [视觉问答](#视觉问答)  
 [图像外推](#图像外推)  
 [光流估计](#光流估计)  
 [视觉定位](#视觉定位)  
 [视觉关系](#视觉关系)  
 [图像增强](#图像增强)  
 [其他](#其他-2)  
 [人脸相关](#人脸相关)  
 [风格迁移](#风格迁移)  
 [人群计数](#人群计数)  
 [轨迹预测](#轨迹预测)  
 [目标检测](#目标检测)  
 [边缘检测](#边缘检测)  
 [边缘计算](#边缘计算)  
 [3D重建](#X1c93a7bf4e1ae71e587568ae7886512738bedde)  
 [ReID](#reid)  
 [行人相关](#行人相关)  
 [行人行为预测](#行人行为预测)  
 [行人检测](#行人检测)  
 [行人属性识别](#行人属性识别)  
 [3D目标检测](#Xa1a3301bccbfe4677fe8ee2f35f74f03a30c7a5)  
 [2D姿态估计](#Xa5fe272a7a1a318436c47bf538495959b9d0239)  
 [3D动作预测](#X405787f33f2d3aeda8d8a9136d5405a84a6d332)  
 [3D姿态估计](#Xd9e7bb27209aa4ee323ba72cd4bbf4cae747dc7)  
 [6D姿态估计](#Xaf104dc90de58abd72b9ddd21d0f11d3ae78929)  
 [Transformer](#transformer)  
 [NeRF、深度学习图形学结合](#nerf深度学习图形学结合)  
 [实例分割](#实例分割)  
 [3D实例分割](#X27e11c0a5a7670934bfc72f49072e120759f4da)  
 [语义分割](#语义分割)  
 [3D语义分割](#X2667005cd29a9f161f2cb1b0523ff6faafdaded)  
 [3D形状识别](#Xec483359703f3752fb46950d07fafa294eb378d)  
 [显著检测](#显著检测)  
 [点云](#点云)  
 [其他](#其他-3)  
 [物体定位](#物体定位-2)  
 [点云去噪](#点云去噪)  
 [点云分割](#点云分割)  
 [点云配准](#点云配准)  
 [点云补全](#点云补全)  
 [目标追踪](#目标追踪)  
 [Video相关](#video相关)  
 [Video动作分类](#video动作分类)  
 [Video分割](#video分割)  
 [Video检测](#video检测)  
 [Video识别](#video识别)  
 [Video插帧](#video插帧)  
 [Video对齐](#video对齐)  
 [Video QA](#video-qa)  
 [Video动作定位](#video动作定位)  
 [Video质量评估](#video质量评估)  
 [Video自然语言查询定位](#video自然语言查询定位)  
 [Video去噪、去模糊](#video去噪去模糊)  
 [Video超分辨率](#video超分辨率)  
 [Video生成](#video生成)  
 [Video异常检测](#video异常检测)  
 [Video inpainting](#video-inpainting)  
 [Video Entailment](#video-entailment)  
 [视频文本检索](#视频文本检索)  
 [视频预测](#视频预测)  
 [主动speaker检测](#主动speaker检测)  
 [视频编码](#视频编码)  
 [视频表征学习](#视频表征学习)  
 [其他](#其他-4)  
 [AI视觉安全](#ai视觉安全)  
 [场景图生成](#场景图生成)  
 [医学相关](#医学相关)  
 [深度估计、预测](#深度估计预测)  
 [手语相关](#手语相关)  
 [强化学习](#强化学习)  
 [联邦学习](#联邦学习)  
 [自动驾驶](#自动驾驶)  
 [Object导航](#object导航)  
 [事件摄像机](#事件摄像机)  
 [人机交互检测](#人机交互检测)  
 [其他](#其他-5)

## 一般的深度学习、机器学习任务

### 网络设计

1. Cherry-Picking Gradients: Learning Low-Rank Embeddings of Visual Data via Differentiable Cross-Approximation 通过可微交叉近似学习视觉数据的低秩嵌入，使得模型可以处理更大、更高分辨率的数据
2. ReCU: Reviving the Dead Weights in Binary Neural Networks 二元神经网络，研究几乎不变的一组权重
3. Aggregation With Feature Detection 改进特征聚合，关注不同特征的语义，追求特征的一致和兼容
4. Dynamic Dual Gating Neural Networks 空间维度估计哪些区域必要，通道维度预测对结果贡献更大的通道，推理时跳过不重要的
5. Learning Specialized Activation Functions With the Piecewise Linear Unit 新的激活函数
6. Adaptive Convolutions With Per-Pixel Dynamic Filter Atom 特征相关网络权值为基础做出的工作，分解filter
7. Causal Attention for Unbiased Visual Recognition 因果注意模块，无监督方式对混淆因素进行注释
8. Refining Activation Downsampling With SoftPool 一种新的下采样方法，可以保留更多的信息
9. FATNN: Fast and Accurate Ternary Neural Networks 三元神经网络
10. Unifying Nonlocal Blocks for Neural Networks 非局部块改进，更鲁棒、灵活地捕获长程依赖
11. Differentiable Dynamic Wirings for Neural Networks 提高模型容量
12. Memory-Augmented Dynamic Neural Relational Inference
13. ShapeConv: Shape-Aware Convolutional Layer for Indoor RGB-D Semantic Segmentation 形状感知卷积层理解深度特征，为RGBD语义分割设计
14. Why Approximate Matrix Square Root Outperforms Accurate SVD in Global Covariance Pooling?
15. Spectral Leakage and Rethinking the Kernel Size in CNNs
16. Neural TMDlayer: Modeling Instantaneous Flow of Features via SDE Generators
17. Poly-NL: Linear Complexity Non-Local Layers With 3rd Order Polynomials
18. FcaNet: Frequency Channel Attention Networks
19. Improving Generalization of Batch Whitening by Convolutional Unit Optimization

### 分类

1. Semantic Diversity Learning for Zero-Shot Multi-Label Classification 端到端图像多标签分类，支持图像和标签的语义多样性
2. Distributional Robustness Loss for Long-Tail Learning 针对长尾分布，鼓励模型学习头类和尾类的高质量表示
3. Learning Rare Category Classifiers on a Tight Labeling Budget 从少量示例中学习罕见类别，而不是尝试平衡数据集
4. Co-Scale Conv-Attentional Image Transformers 改进注意力机制
5. Semi-Supervised Learning of Visual Features by Non-Parametrically Predicting View Assignments With Support Samples 确保同一未标记实例的不同视图分配相似的为标签
6. Webly Supervised Fine-Grained Recognition: Benchmark Datasets and an Approach 提供了数据集
7. Transformer-Based Dual Relation Graph for Multi-Label Image Recognition 识别一幅图像中多个物体
8. On the Importance of Distractors for Few-Shot Classification 基于对比学习的改进微调方法，加入干扰物提高泛化能力
9. ACE: Ally Complementary Experts for Solving Long-Tailed Recognition in One-Shot 针对单阶段长尾方法的“seesaw”问题进行改进
10. Partner-Assisted Learning for Few-Shot Image Classification 着重训练策略的设计
11. Influence-Balanced Loss for Imbalanced Visual Classification 平衡训练方法来解决不平衡数据学习的问题
12. Benchmark Platform for Ultra-Fine-Grained Visual Categorization Beyond Human Performance 超细粒度图像分类数据集
13. A Unified Objective for Novel Class Discovery NCD问题：利用包含不同但相关类的标记集的先验知识，在未标记集中推断新的对象类别
14. A Multi-Mode Modulator for Multi-Domain Few-Shot Classification
15. Conditional Variational Capsule Network for Open Set Recognition 开集（标签空间中包含/未包含的样本称为已知/未知）识别
16. Variational Feature Disentangling for Fine-Grained Few-Shot Classification
17. Group-Wise Inhibition Based Feature Regularization for Robust Classification
18. Searching for Robustness: Loss Learning for Noisy Classification Tasks
19. Relational Embedding for Few-Shot Classification
20. Residual Attention: A Simple but Effective Method for Multi-Label Recognition 多标签分类，简单但有效

### 聚类

1. Learning Hierarchical Graph Neural Networks for Image Clustering 不是完全无监督的层次GNN用于聚类
2. One-Pass Multi-View Clustering for Large-Scale Data 多视图聚类算法？
3. Multi-VAE: Learning Disentangled View-Common and View-Peculiar Visual Representations for Multi-View Clustering
4. Clustering by Maximizing Mutual Information Across Views 图像聚类框架

### 无监督/自监督

1. Focus on the Positives: Self-Supervised Learning for Biodiversity Monitoring
2. ISD: Self-Supervised Learning by Iterative Similarity Distillation 改进对比学习中的相似性标记问题
3. On Feature Decorrelation in Self-Supervised Learning 特征去相关
4. Rethinking 360deg Image Visual Attention Modelling With Unsupervised Learning 360图像学习
5. Instance Similarity Learning for Unsupervised Feature Representation 新方法学习实例间的语义相似度
6. Geography-Aware Self-Supervised Learning 地理定位数据集上的对比学习应用
7. Co2L: Contrastive Continual Learning 对比学习加自监督蒸馏的持续学习
8. Self-Supervised Representation Learning From Flow Equivariance 流变换自监督表征学习，视频学的表示应用于图像下游任务
9. Efficient Visual Pretraining With Contrastive Detection 减少自监督预训练需要的计算量
10. Improving Contrastive Learning by Visualizing Feature Transformation 不同于数据增强的特征级别数据操作，改进对比学习
11. CDS: Cross-Domain Self-Supervised Pre-Training 跨域自监督预训练
12. Solving Inefficiency of Self-Supervised Representation Learning 新的自监督学习框架，提高自监督对比学习的学习效率
13. **Contrastive Multimodal Fusion With TupleInfoNCE** 自监督多模态
14. Preservational Learning Improves Self-Supervised Medical Image Models by Reconstructing Diverse Contexts

### 半监督

1. CoMatch: Semi-Supervised Learning With Contrastive Graph Regularization 统一了主流方法并解决一些局限性

### 对抗

1. Learning Bias-Invariant Representation by Cross-Sample Mutual Information Minimization 对抗性模型去偏
2. Feature Importance-Aware Transferable Adversarial Attacks 特征重要性攻击
3. Meta Gradient Adversarial Attack 即插即用
4. Low Curvature Activations Reduce Overfitting in Adversarial Training 激活函数的选择影响对抗性训练的过拟合现象
5. Batch Normalization Increases Adversarial Vulnerability and Decreases Adversarial Transferability: A Non-Robust Feature Perspective 研究Batch Normalization 认为其导致模型学到了不鲁棒的特征
6. Relating Adversarially Robust Generalization to Flat Minima 针对对抗性训练过拟合，对AT过程做出改进
7. Towards Understanding the Generative Capability of Adversarially Robust Classifiers 研究对抗鲁棒分类器的生成性能的工作，并针对这点提出一个更好的 对抗训练方法
8. Dynamic Divide-and-Conquer Adversarial Training for Robust Semantic Segmentation 研究对抗性训练在分割中应用的工作
9. Adversarial Example Detection Using Latent Neighborhood Graph 基于图的对抗性检测方法
10. Learnable Boundary Guided Adversarial Training
11. AGKD-BML: Defense Against Adversarial Attack by Attention Guided Knowledge Distillation and Bi-Directional Metric Learning
12. Reliably Fast Adversarial Training via Latent Adversarial Perturbation 多步方法计算开销很高，单步性能不可靠，提出改进的单步潜在对抗训练方法

### 网络加速

1. RDI-Net: Relational Dynamic Inference Networks 改进以往方法的router，并增加一个新的用来配合的模块
2. Towards Memory-Efficient Neural Networks via Multi-Level In Situ Generation 硬件结合加速

### 模型剪枝

1. ResRep: Lossless CNN Pruning via Decoupling Remembering and Forgetting
2. GDP: Stabilized Neural Network Pruning via Gates With Differentiable Polarization

### 知识蒸馏

1. Densely Guided Knowledge Distillation Using Multiple Teacher Assistants 多teacher引导，逐渐减小模型尺寸，弥补teacher和student的差距
2. G-DetKD: Towards General Distillation Framework for Object Detectors via Contrastive and Semantic-Guided Feature Imitation
3. Revisiting Adversarial Robustness Distillation: Robust Soft Labels Make Student Better
4. Lipschitz Continuity Guided Knowledge Distillation 更合理的数学模型指导知识蒸馏
5. Distilling Virtual Examples for Long-Tailed Recognition 从知识蒸馏角度解决长尾视觉识别问题
6. Compressing Visual-Linguistic Model via Knowledge Distillation 视觉知识蒸馏压缩语言表征的预训练模型
7. Self-Knowledge Distillation With Progressive Refinement of Targets 知识蒸馏增加泛化
8. Distilling Global and Local Logits With Densely Connected Relations 指出gap计算logits的局限性，做了一些改进
9. Multimodal Knowledge Expansion
10. Student Customized Knowledge Distillation: Bridging the Gap Between Student and Teacher
11. Channel-Wise Knowledge Distillation for Dense Prediction 最小化两个网络的通道概率图之间的KL散度

### Loss

1. ARAPReg: An As-Rigid-As Possible Regularization Loss for Learning Deformable Shape Generators
2. Orthogonal Projection Loss 在特征空间中具有正交性，在特征空间中直接实现类间分离和类内聚类
3. Bias Loss for Mobile Neural Networks 改造标准的交叉熵使紧凑的模型学习有价值的数据点，从而有特征多样性
4. Do Different Deep Metric Learning Losses Lead to Similar Learned Features? 研究不同loss对于模型的影响
5. Learning Multiple Pixelwise Tasks Based on Loss Scale Balancing 新的损失加权算法，用于像素视觉任务的多任务学习
6. Towards Learning Spatially Discriminative Feature Representations 新的CAM-loss，还有一种根据loss提出的知识蒸馏方法

### 模型量化

1. Improving Low-Precision Network Quantization via Bin Regularization 权值正则化算法
2. Bit-Mixer: Mixed-Precision Networks With Runtime Bit-Width Selection
3. Towards Mixed-Precision Quantization of Neural Networks via Constrained Optimization

### 迁移学习

1. Few-Shot Image Classification: Just Use a Library of Pre-Trained Feature Extractors and a Simple Classifier
2. Z-Score Normalization, Hubness, and Few-Shot Learning 针对hubness问题提出的使用Z-Score
3. Generalized and Incremental Few-Shot Learning by Explicit Learning and Calibration Without Forgetting 三阶段框架，校准分类器防止遗忘
4. DeepGaze IIE: Calibrated Prediction in and Out-of-Domain for State-of-the-Art Saliency Modeling
5. Pseudo-Loss Confidence Metric for Semi-Supervised Few-Shot Learning 半监督few-shot，统一置信度估计方法
6. Field-Guide-Inspired Zero-Shot Learning 研究zero-shot的注释方法
7. Semantics Disentangling for Generalized Zero-Shot Learning GZSL（假设某些类是不可见的情况下对样本进行分类）与ZSL的区别暂时不清楚
8. **Hierarchical Graph Attention Network for Few-Shot Visual-Semantic Learning few-shot视觉语义（图像-文本样本多模态）学习**
9. Learning To Hallucinate Examples From Extrinsic and Intrinsic Supervision 从学习Hallucinate Examples这个角度去改进FSL的表现
10. Discriminative Region-Based Multi-Label Zero-Shot Learning 提出了一种基于区域的保持可分辨性的方法
11. Curvature Generation in Curved Spaces for Few-Shot Learning
12. FREE: Feature Refinement for Generalized Zero-Shot Learning 特征细化的GZSL，又提出了两种损失函数指导特征细化模块学习类和语义相关表示
13. Task-Aware Part Mining Network for Few-Shot Learning
14. Fast and Efficient DNN Deployment via Deep Gaussian Transfer Learning
15. TransferI2I: Transfer Learning for Image-to-Image Translation From Small Datasets I2I转换的迁移学习框架

### 可解释性

1. Finding Representative Interpretations on Convolutional Neural Networks 无监督方法解释CNN决策逻辑
2. Explanations for Occluded Images
3. Architecture Disentanglement for Deep Neural Networks 将DNN分解为子架构，形成描述推理过程的信息流
4. Explaining Local, Global, and Higher-Order Interactions in Deep Learning 解释NN跨多对象的关系推理
5. Beyond Trivial Counterfactual Explanations With Diverse Valuable Explanations
6. Generating Attribution Maps With Disentangled Masked Backpropagation 属性图可视化任务，计算每个像素对网络输出的贡献度
7. SCOUTER: Slot Attention-Based Classifier for Explainable Image Recognition 新的分类器，用于透明而准确的分类
8. Statistically Consistent Saliency Estimation
9. Generic Attention-Model Explainability for Interpreting Bi-Modal and Encoder-Decoder Transformers

### 持续学习

1. Online Continual Learning With Natural Distribution Shifts: An Empirical Study With Visual Data 在线持续学习（在多个任务中学习和保留知识）
2. Rehearsal Revealed: The Limits and Merits of Revisiting Samples in Continual Learning 解释研究持续学习中的rehearsal方法
3. RECALL: Replay-Based Continual Learning in Semantic Segmentation 持续学习在语义分割中的应用

### 主动学习

1. Contrastive Coding for Active Learning Under Class Distribution Mismatch 基于类分布不匹配的问题，提出一种基于对比编码的AL框架
2. Semi-Supervised Active Learning for Semi-Supervised Models: Exploit Adversarial Examples With Graph-Based Virtual Labels 主动学习和半监督结合框架的改进

### 元学习

1. A Lazy Approach to Long-Horizon Gradient-Based Meta-Learning 基于梯度的元学习
2. Meta-Learning With Task-Adaptive Loss Function for Few-Shot Learning 改进的用于FSL的元学习框架
3. Meta-Baseline: Exploring Simple Meta-Learning for Few-Shot Learning

### 训练

1. DecentLaM: Decentralized Momentum SGD for Large-Batch Deep Training 适用于分布式训练的新的SGD

### 深度度量学习

1. LoOp: Looking for Optimal Hard Negative Embeddings for Deep Metric Learning 深度度量学习，提出新的方法
2. Deep Relational Metric Learning 用于图像聚类和检索的深度关系度量学习
3. Learning Better Visual Data Similarities via New Grouplet Non-Euclidean Embedding

### 其他

1. Continual Learning on Noisy Data Streams via Self-Purified Replay 通过一些观察设计了两个组件，得到比SOTA持续学习方法和有噪声的标签学习方法组合后更好的效果
2. NGC: A Unified Framework for Learning With Open-World Noisy Data 分布外（OOD）噪声清除
3. Self-Born Wiring for Neural Trees 神经网络+决策树，新框架
4. Multi-Task Self-Training for Learning General Representations 通用模型学习 多任务自我训练框架 Teacher & Student
5. Predicting With Confidence on Unseen Distributions
6. Discover the Unknown Biased Attribute of an Image Classifier 新问题，发现分类器对输入图像未知的偏差属性
7. Influence Selection for Active Learning 新的主动学习样本选择策略，模型和任务无关
8. Impact of Aliasing on Generalization in Deep Convolutional Networks 研究aliasing对泛化的影响
9. SS-IL: Separated Softmax for Incremental Learning 类增量学习，分离softmax结合任务智能知识提取层
10. Always Be Dreaming: A New Approach for Data-Free Class-Incremental Learning
11. Modulated Periodic Activations for Generalizable Local Functional Representations
12. Learning With Noisy Labels via Sparse Regularization 在有噪声的标签学习下增加鲁棒性
13. Meta Learning on a Sequence of Imbalanced Domains With Difficulty Awareness 假设一个元学习更困难的环境，提出新的方法
14. MixMix: All You Need for Data-Free Compression Are Feature and Data Mixing 原数据被保护的情况下， 研究使用预训练模型生成数据训练
15. MixMo: Mixing Multiple Inputs for Multiple Outputs via Deep Subnetworks 研究多输入多输出架构，更好地混合多个输入
16. The Many Faces of Robustness: A Critical Analysis of Out-of-Distribution Generalization 研究分布外鲁棒性，得出与之前工作相反的结论，还提出几个新的数据集
17. Exploring Relational Context for Multi-Task Dense Prediction 多任务环境中根据任务间的关系捕获跨任务上下文并优化每个任务
18. Robustness via Cross-Domain Ensembles 使神经网络预测对训练数据分布的变化具有鲁棒性的研究
19. Low-Shot Validation: Active Importance Sampling for Estimating Classifier Performance on Rare Categories 标记数据有限的情况下，减少用于验证的注释成本，提出一种统计验证法准确估计稀少类别的F分数
20. Learning Compatible Embeddings 模型兼容性相关工作
21. JEM++: Improved Techniques for Training JEM
22. Adaptive Curriculum Learning CL：先学习较容易的概念，后关注较难的概念。示例难度根据模型当前状态自适应调整
23. Learning With Privileged Tasks 多目标多任务学习，新算法仲裁目标任务和特权任务的学习
24. Global Pooling, More Than Meets the Eye: Position Information Is Encoded Channel-Wise in CNNs
25. Testing Using Privileged Information by Adapting Features With Statistical Dependence 测试时用额外的特征来改进所做的预测
26. Ranking Models in Unlabeled New Environments 未标记新环境下对现有的模型进行排序评估
27. Kernel Methods in Hyperbolic Spaces
28. A Machine Teaching Framework for Scalable Recognition
29. DisUnknown: Distilling Unknown Factors for Disentanglement Learning 数据解纠缠学习

## 脉冲神经网络

1. HIRE-SNN: Harnessing the Inherent Robustness of Energy-Efficient Deep Spiking Neural Networks by Training With Crafted Input Noise 新的训练脉冲神经网络的方法，增加鲁棒性
2. Event Stream Super-Resolution via Spatiotemporal Constraint Learning 脉冲神经网络的事件流摄像机超分辨率端到端框架
3. DCT-SNN: Using DCT To Distribute Spatial Information Over Time for Low-Latency Spiking Neural Networks 利用离散余弦变换加快推理

## 图神经网络

1. Meta-Aggregator: Learning To Aggregate for 1-Bit Graph Neural Networks
2. Consistency-Aware Graph Network for Human Interaction Understanding 图神经网络人类交互理解
3. Hypergraph Neural Networks for Hypergraph Matching 超图匹配的第一个超图神经网络解决方案
4. Learning To Bundle-Adjust: A Graph Network Approach to Faster Optimization of Bundle Adjustment for Vehicular SLAM 图网络加速BA

## 数据增广

1. Better Aggregation in Test-Time Augmentation 研究测试时的数据增广
2. Towards Rotation Invariance in Object Detection 新的旋转数据增强方法
3. DivAug: Plug-In Automated Data Augmentation With Explicit Diversity Maximization 自动学习的增广策略研究
4. Context Decoupling Augmentation for Weakly Supervised Semantic Segmentation
5. Point Cloud Augmentation With Weighted Local Transformations 点云数据增广
6. Direct Differentiable Augmentation Search
7. TrivialAugment: Tuning-Free Yet State-of-the-Art Data Augmentation 自动增强方法的研究

## NAS

1. Pi-NAS: Improving Neural Architecture Search by Reducing Supernet Training Consistency Shift
2. FairNAS: Rethinking Evaluation Fairness of Weight Sharing Neural Architecture Search
3. GLiT: Neural Architecture Search for Global and Local Image Transformer 使用NAS找到一个更好的transformer架构用于图像识别
4. Achieving On-Mobile Real-Time Super-Resolution With Neural Architecture and Pruning Search NAS和剪枝实现移动设备上的单图像超分辨率
5. BN-NAS: Neural Architecture Search With Batch Normalization 加快评估速度，超网训练阶段只训练BN参数
6. Homogeneous Architecture Augmentation for Neural Predictor 改进神经预测器增加NAS的速度和表现
7. Neural Architecture Search for Joint Human Parsing and Pose Estimation 使用NAS搜索两个密切相关任务的模型
8. EC-DARTS: Inducing Equalized and Consistent Optimization Into DARTS NAS的变种方法的改进
9. Distilling Optimal Neural Networks: Rapid Search in Diverse Spaces 快速、可扩展和多样化NAS新pipeline
10. Not All Operations Contribute Equally: Hierarchical Operation-Adaptive Predictor for Neural Architecture Search 基于图的预测器对所有操作都一视同仁，导致单元结构的拓扑知识存在偏差
11. Learning Latent Architectural Distribution in Differentiable Neural Architecture Search via Variational Information Maximization 利用一个简单但有效的卷积神经网络多潜在表示进行建模
12. Zen-NAS: A Zero-Shot NAS for High-Performance Image Recognition 使用新的指标对搜索的结构进行排序
13. CM-NAS: Cross-Modality Neural Architecture Search for Visible-Infrared Person Re-Identification
14. BossNAS: Exploring Hybrid CNN-Transformers With Block-Wisely Self-Supervised Neural Architecture Search
15. NASOA: Towards Faster Task-Oriented Online Fine-Tuning With a Zoo of Models 联合NAS和自适应框架

### 视觉表征学习

1. A Broad Study on the Transferability of Visual Representations With Contrastive Learning 研究对比学习向不同领域的迁移能力
2. Learning From Noisy Data With Robust Representation Learning 关注三种常见噪声，不止标签噪声
3. Vi2CLR: Video and Image for Visual Contrastive Learning of Representation 自监督同时理解图像和视频
4. Contrasting Contrastive Self-Supervised Representation Learning Pipelines 大量实验分析对比学习对视觉表征学习的作用
5. Temporal Knowledge Consistency for Unsupervised Visual Representation Learning 观察到在训练的不同阶段同一实例教师给出的监督会相差很大，导致问题，时间知识一致性改进
6. With a Little Help From My Friends: Nearest-Neighbor Contrastive Learning of Visual Representations 近邻对比学习
7. Pri3D: Can 3D Priors Help 2D Representation Learning? 3D先验编码为2D表征
8. Weakly Supervised Representation Learning With Coarse Labels 粗粒度标签进行细粒度学习
9. LocTex: Learning Data-Efficient Visual Representations From Localized Textual Supervision 使用低成本localized文本注释（标题和同步鼠标移动手势）训练，注释工作量减少，性能更好。
10. Concept Generalization in Visual Representation Learning 概念泛化中可见和不可见概念的语义关系会影响泛化性能
11. Contrastive Learning of Image Representations With Cross-Video Cycle-Consistency 新的对比学习方法，利用循环一致性进行表征学习
12. Exploring Visual Engagement Signals for Representation Learning 利用视觉参与信号（社交媒体平台的视觉参与包括与照片帖子的互动，包括评论、分享和喜欢）监督的表征学习
13. Switchable K-Class Hyperplanes for Noise-Robust Representation Learning
14. Region Similarity Representation Learning

### DA

1. S3VAADA: Submodular Subset Selection for Virtual Adversarial Active Domain Adaptation
2. Zero-Shot Day-Night Domain Adaptation With a Physics Prior
3. Active Universal Domain Adaptation 引入主动学习
4. Semantic Concentration for Domain Adaptation 缓解不相关的语义信息对特征对齐造成的负面影响
5. Re-Energizing Domain Discriminator With Sample Relabeling for Adversarial Domain Adaptation
6. Transporting Causal Mechanisms for Unsupervised Domain Adaptation 因果机制解决语义损失，鼓励模型学习跨域共同特征
7. Active Domain Adaptation via Clustering Uncertainty-Weighted Embeddings 主动学习+DA，聚类不确定性加权？
8. Recursively Conditional Gaussian for Ordinal Unsupervised Domain Adaptation 无监督DA，针对离散连续标签替换了iid先验
9. OVANet: One-vs-All Network for Universal Domain Adaptation 处理两个数据集之间的域转移和类别转移
10. Collaborative Optimization and Aggregation for Decentralized Domain Generalization and Adaptation
11. **Collaborative Learning With Disentangled Features for Zero-Shot Domain Adaptation** 新的概念zero-shot DA，对抗式学习对领域不变的语义特征和任务不变的领域特征进行协作训练，从而学习任务不可知的领域转移
12. Partial Video Domain Adaptation With Partial Adversarial Temporal Attentive Network 第一次接触的概念Partial DA
13. Information-Theoretic Regularization for Multi-Source Domain Adaptation 解决多个领域鉴别器对MDA的不利影响
14. **Sparse-to-Dense Feature Matching: Intra and Inter Domain Cross-Modal Learning in Domain Adaptation for 3D Semantic Segmentation 跨域跨模态学习**
15. Multi-Target Adversarial Frameworks for Domain Adaptation in Semantic Segmentation UDA应用在语义分割
16. Adversarial Robustness for Unsupervised Domain Adaptation 研究UDA的对抗鲁棒性
17. Gradient Distribution Alignment Certificates Better Adversarial Domain Adaptation 约束两个域的特征梯度具有相似的分布
18. SPG: Unsupervised Domain Adaptation for 3D Object Detection via Semantic Point Generation 基于3D检测的UDA
19. Adaptive Adversarial Network for Source-Free Domain Adaptation
20. Geometric Unsupervised Domain Adaptation for Semantic Segmentation 自监督单目深度估计作为代理任务
21. Unsupervised Real-World Super-Resolution: A Domain Adaptation Perspective
22. T-SVDNet: Exploring High-Order Prototypical Correlations for Multi-Source Domain Adaptation
23. SENTRY: Selective Entropy Optimization via Committee Consistency for Unsupervised Domain Adaptation
24. Self-Supervised Domain Adaptation for Forgery Localization of JPEG Compressed Images
25. Towards Discriminative Representation Learning for Unsupervised Person Re-Identification 针对以前UDA框架的一些局限做了一些改进
26. ECACL: A Holistic Framework for Semi-Supervised Domain Adaptation 半监督DA，利用源域中的标记样本，在目标域中使用未标记样本和少量标记样本学习性能良好的模型
27. PIT: Position-Invariant Transform for Cross-FoV Domain Adaptation 注意到FoV带来的域gap
28. Generalized Source-Free Domain Adaptation

## DG

1. Domain Generalization via Gradient Surgery
2. **SelfReg: Self-Supervised Contrastive Regularization for Domain Generalization 考虑到负数据质量对于对比学习的影响，研究只使用正数据的方法**
3. A Style and Semantic Memory Mechanism for Domain Generalization 主流方法有限考虑跨域语义不变，低估域内的风格不变
4. Confidence Calibration for Domain Generalization Under Covariate Shift

## 生成式

1. MeshTalk: 3D Face Animation From Speech Using Cross-Modality Disentanglement 基于语音的3D人脸动画生成
2. AD-NeRF: Audio Driven Neural Radiance Fields for Talking Head Synthesis 音频序列生成高保真说话人头
3. Explain Me the Painting: Multi-Topic Knowledgeable Art Description Generation 艺术作品生成信息性描述
4. Full-Body Motion From a Single Head-Mounted Device: Generating SMPL Poses From Partial Observations 利用头戴式和可穿戴设备的信号流构建用户的表示
5. FACIAL: Synthesizing Dynamic Talking Face With Implicit Attribute Learning
6. Handwriting Transformers 手写文本生成
7. Learning High-Fidelity Face Texture Completion Without Complete Face Texture 在不使用任何完整纹理的情况下，学习在单个人脸图像中补全不可见纹理
8. Neural Strokes: Stylized Line Drawing of 3D Shapes 3D形状生成线条图
9. Attack-Guided Perceptual Data Generation for Real-World Re-Identification

### 图像生成

1. Geometry-Free View Synthesis: Transformers and No 3D Priors 给定一张图和一个期望的视角，根据视角生成新的图
2. Image Synthesis via Semantic Composition 根据区域间的外观相关性建立依赖关系
3. Unsupervised Image Generation With Infinite Generative Adversarial Networks 根据GAN已有的问题进行改进，构造潜在空间、避免模态坍塌
4. Image Synthesis From Layout With Locality-Aware Mask Adaption 在有一组带有对象类别的边界框的情况下合成图像
5. CanvasVAE: Learning To Generate Vector Graphic Documents 矢量图形文档生成
6. Dual Contrastive Loss and Attention for GANs 使用双重对比损失和改进注意力机制来扩展图像生成的边界
7. ILVR: Conditioning Method for Denoising Diffusion Probabilistic Models 指导DDPM可控地生成高质量图像
8. Rethinking Noise Synthesis and Modeling in Raw Denoising 缺乏大规模的真实原始图像去噪数据集，介绍了一种从传感器真实噪声中直接采样合成噪声的新方法，DNN合成的噪声图像有问题
9. Semi-Supervised Single-Stage Controllable GANs for Conditional Fine-Grained Image Generation 减轻模型设计和训练的复杂性，单阶段可控GAN，三个变量进行生成建模
10. Learning Realistic Human Reposing Using Cyclic Self-Supervision With 3D Shape, Pose, and Appearance Consistency 一张图片合成一个人的新姿势
11. Toward Spatially Unbiased Generative Models 显式位置编码，学习空间无偏生成器
12. Orthogonal Jacobian Regularization for Unsupervised Disentanglement in Image Generation
13. Teacher-Student Adversarial Depth Hallucination To Improve Face Recognition RGB图像生成深度图像，在不可见的数据集中很好地泛化，人脸识别下游任务性能更好
14. Dressing in Order: Recurrent Person Image Generation for Pose Transfer, Virtual Try-On and Outfit Editing 2D人体姿势转化、虚拟试穿等
15. Physics-Based Differentiable Depth Sensor Simulation
16. PIRenderer: Controllable Portrait Image Generation via Semantic Neural Rendering
17. Image Shape Manipulation From a Single Augmented Training Sample 基于单个图像的条件图像处理生成，广泛的增强是实现单图像训练的关键
18. GANcraft: Unsupervised 3D Neural Rendering of Minecraft Worlds
19. Diagonal Attention and Style-Based GAN for Content-Style Disentanglement in Image Generation and Translation 将内容空间和风格空间解纠缠进行单独控制
20. Aligning Latent and Image Spaces To Connect the Unconnectable 生成无限多个内容复杂的高分辨率图像并提出新的数据集

### 3D生成

1. SurfGen: Adversarial 3D Shape Synthesis With Explicit Surface Discriminators
2. DeepCAD: A Deep Generative Network for Computer-Aided Design Models 建立了一个数据集，将3D形状生成表示为一系列的CAD操作
3. Indoor Scene Generation From a Collection of Semantic-Segmented Depth Images 利用从不同未知场景中获得的语义分割深度图集合进行3D室内场景生成
4. 3D Shape Generation and Completion Through Point-Voxel Diffusion 无条件形状生成和有条件的多模态形状补全的统一
5. imGHUM: Implicit Generative Models of 3D Human Shape and Articulated Pose 3D人体形状和关节姿势的整体生成模型
6. LSD-StructureNet: Modeling Levels of Structural Detail in 3D Part Hierarchies 可以重新生成输出层次结构中任意位置的部件
7. Neural-GIF: Neural Generalized Implicit Functions for Animating People in Clothing 输入骨架pose，生成带有衣着的人体3D模型
8. 3D-FRONT: 3D Furnished Rooms With layOuts and semaNTics 数据集
9. Physics-Based Human Motion Estimation and Synthesis From Videos 从单目RGB视频生成合理的人体运动模型
10. Scene Synthesis via Uncertainty-Driven Attribute Synchronization
11. Learning Generative Models of Textured 3D Meshes From Real-World Images 之前的工作依赖姿势估计注释，进行改进
12. Action-Conditioned 3D Human Motion Synthesis With Transformer VAE 动作条件下生成真实和多样的人体运动序列
13. 3DStyleNet: Creating 3D Shapes With Geometric and Texture Style Variations
14. Deep Hybrid Self-Prior for Full 3D Mesh Generation
15. Learned Spatial Representations for Few-Shot Talking-Head Synthesis
16. Stochastic Scene-Aware Motion Prediction 3D场景中生成3D人体模型
17. Graph-to-3D: End-to-End Generation and Manipulation of 3D Scenes Using Scene Graphs
18. DeePSD: Automatic Deep Skinning and Pose Space Deformation for 3D Garment Animation

### GAN

1. Sketch Your Own GAN
2. LatentCLR: A Contrastive Learning Approach for Unsupervised Discovery of Interpretable Directions 在预训练好的GAN的潜在空间中找到可解释的方向
3. GAN-Control: Explicitly Controllable GANs 训练显式可控的生成面部图像的GAN
4. Omni-GAN: On the Secrets of cGANs and Beyond 改进cGAN，保证鉴别器在感知概念时受到较强的监督
5. DAE-GAN: Dynamic Aspect-Aware GAN for Text-to-Image Synthesis 更好地利用纵横比信息
6. Detail Me More: Improving GAN's Photo-Realism of Complex Scenes 在适当的区域协调鉴别器
7. Intrinsic-Extrinsic Preserved GANs for Unsupervised 3D Pose Transfer 形状和姿势进行信息保持，新的GAN结构进行3D姿势转换
8. DRB-GAN: A Dynamic ResBlock Generative Adversarial Network for Artistic Style Transfer 利用GAN进行风格迁移，可能是新的style code建模方式？
9. OpenGAN: Open-Set Recognition via Open Data Generation 区分K闭集类之外的开集数据
10. Online Multi-Granularity Distillation for GAN Compression 在线多粒度蒸馏压缩GAN
11. Toward a Visual Concept Vocabulary for GAN Latent Space
12. SurfaceNet: Adversarial SVBRDF Estimation From a Single Image
13. Gradient Normalization for Generative Adversarial Networks 解决GAN训练不稳
14. EigenGAN: Layer-Wise Eigen-Learning for GANs
15. HeadGAN: One-Shot Neural Head Synthesis and Editing
16. Explaining in Style: Training a GAN To Explain a Classifier in StyleSpace 一个新的训练GAN的方法来理解分类器
17. StyleCLIP: Text-Driven Manipulation of StyleGAN Imagery
18. When Do GANs Replicate? On the Choice of Dataset Size
19. InSeGAN: A Generative Approach to Segmenting Identical Instances in Depth Images
20. WarpedGANSpace: Finding Non-Linear RBF Paths in GAN Latent Space
21. From Continuity to Editability: Inverting GANs With Consecutive Images
22. Towards Discovery and Attribution of Open-World GAN Generated Images 识别图像并归因于GAN，迭代算法识别训练期间不可见的GAN生成的图像

## 动作识别

1. AdaSGN: Adapting Joint Number and Model Size for Efficient Skeleton-Based Action Recognition 通过自适应地 controll the input number of the joints of the skeleton on-the-fly 的方式来减少计算代价并且可以自适应地选择模型大小，加快动作识别
2. Selective Feature Compression for Efficient Activity Recognition Inference 在时空维度上压缩也正，加快推理
3. OadTR: Online Action Detection With Transformers 使用Transformer代替RNN
4. Self-Supervised 3D Skeleton Action Representation Learning With Motion Consistency and Continuity 自监督动作识别
5. Skeleton Cloud Colorization for Unsupervised 3D Action Representation Learning 无监督表示学习
6. Channel-Wise Topology Refinement Graph Convolution for Skeleton-Based Action Recognition 动态学习不同拓扑，有效聚合不同通道中的关节特征
7. GroupFormer: Group Activity Recognition With Clustered Spatial-Temporal Transformer 联合捕获时空上下文信息，有效增强个体和群体表示
8. Spatio-Temporal Dynamic Inference Network for Group Activity Recognition
9. Else-Net: Elastic Semantic Network for Continual Action Recognition From Skeleton Data 从连续的skeleton数据流中学习一个随时间变化的识别模型
10. Geometric Deep Neural Network Using Rigid and Non-Rigid Transformations for Human Action Recognition

## 图像理解

### 图像去噪、去光等

1. C2N: Practical Generative Noise Modeling for Real-World Denoising 提出一个生成图像噪音的框架来更好地帮助进行图像去噪
2. Light Source Guided Single-Image Flare Removal From Unpaired Data 提出了一个框架和新的数据集
3. How To Train Neural Networks for Flare Removal
4. Location-Aware Single Image Reflection Removal 基于位置感知的单图像反射去除，有模块回归概率置信图
5. Single Image Defocus Deblurring Using Kernel-Sharing Parallel Atrous Convolutions 图像离焦去模糊
6. Improving De-Raining Generalization via Neural Reorganization 去雨，将多个合成数据集的知识积累到单个网络参数中，使用了知识蒸馏等方法来设计
7. Rethinking Deep Image Prior for Denoising 深度图像先验去噪
8. MBA-VO: Motion Blur Aware Visual Odometry 补偿运动模糊，提高VO鲁棒性，提供一个新的数据集
9. Towards Flexible Blind JPEG Artifacts Removal 预测可调节的质量因子控制伪影去除和细节保留之间的权衡
10. Learning To Reduce Defocus Blur by Realistically Modeling Dual-Pixel Data 提出生成的DP数据的过程来解决数据捕获困难，并证明对于去模糊有用
11. Learning Dual Priors for JPEG Compression Artifacts Removal 可解释性更强，学习像素级回归先验和语义判别先验
12. Hyperspectral Image Denoising With Realistic Data 之前大多数方法都是基于合成数据集，研究如何生成真实的数据集来学习和评估HSI去噪
13. Learning To Remove Refractive Distortions From Underwater Images 去除水下图像的折射失真
14. Defocus Map Estimation and Deblurring From a Single Dual-Pixel Image 单个双像素图作为输入，同时估计图像散焦图并恢复全聚焦图像
15. End-to-End Unsupervised Document Image Blind Denoising
16. Let's See Clearly: Contaminant Artifact Removal for Moving Cameras 移动相机的伪影去除
17. Unpaired Learning for Deep Image Deraining With Rain Direction Regularizer 图像去雨
18. Self-Supervised Image Prior Learning With GMM From a Single Noisy Image GMM自监督学习，从单个噪声图像实现噪声水平估计和图像先验学习
19. V-DESIRR: Very Fast Deep Embedded Single Image Reflection Removal
20. Rethinking Coarse-To-Fine Approach in Single Image Deblurring
21. Structure-Preserving Deraining With Residue Channel Prior Guidance 发现图像的RCP比雨天图像包含更准确的结构信息

### 图像检索

1. Ask&Confirm: Active Detail Enriching for Cross-Modal Retrieval With Partial Query 强化学习、弱监督，提出新检索框架，交互更友好
2. DOLG: Single-Stage Image Retrieval With Deep Orthogonal Fusion of Local and Global Features 提出了一种深度正交局部与全局(DOLG)信息融合框架用于端到端图像检索
3. Instance-Level Image Retrieval Using Reranking Transformers 根据某个物体实例进行图像检索
4. Image Retrieval on Real-Life Images With Pre-Trained Vision-and-Language Models 扩展了组合图像检索任务，提出了一个数据集
5. Self-Supervised Product Quantization for Deep Unsupervised Image Retrieval 交叉量化对比学习策略
6. Learning Attribute-Driven Disentangled Representations for Interactive Fashion Retrieval 时装检索，利用视觉属性语义训练，学习解纠缠表示
7. Telling the What While Pointing to the Where: Multimodal Queries for Image Retrieval 不只文本，多模态查询图像检索
8. ASMR: Learning Attribute-Based Person Search With Adaptive Semantic Margin Regularizer 查找与作为查询的一组文本属性最匹配的人物图像
9. Universal Cross-Domain Retrieval: Generalizing Across Classes and Domains
10. Learning Deep Local Features With Multiple Dynamic Attentions for Large-Scale Image Retrieval
11. ALADIN: All Layer Adaptive Instance Normalization for Fine-Grained Style Similarity

### 图像变化理解

1. Describing and Localizing Multiple Changes With Transformers 数据集论文，套transformer，多变化捕捉并用简洁的标题描述SOTA
2. Viewpoint-Agnostic Change Captioning With Cycle Consistency 过滤仅仅是改变相机视角的变化
3. Change Is Everywhere: Single-Temporal Supervised Object Change Detection in Remote Sensing Imagery 遥感图像变化检测，利用对象的变化在未配对的图像作为监督信号

### 图像超分辨率

1. Attention-Based Multi-Reference Learning for Image Super-Resolution
2. Learning a Single Network for Scale-Arbitrary Super-Resolution 任意尺度图像SR
3. Fourier Space Losses for Efficient Perceptual Image Super-Resolution 提出新的Loss，在模型规模不大的情况下提供更高效的感知
4. Designing a Practical Degradation Model for Deep Blind Image Super-Resolution 设计退化模型来提高单幅图像超分的表现
5. Dynamic High-Pass Filtering and Multi-Spectral Attention for Image Super-Resolution 根据CNN喜欢学习低频信号做出的改进
6. Context Reasoning Attention Network for Image Super-Resolution 考虑到全局上下文是任务关键，但CNN是提取局部模式的，做出改进
7. Mutual Affine Network for Spatially Variant Kernel Estimation in Blind Image Super-Resolution 认为现有方法的假设不适应真实图像，进行了改进
8. Benchmarking Ultra-High-Definition Image Super-Resolution UHD图像超分辨率数据集和baseline
9. Lucas-Kanade Reloaded: End-to-End Super-Resolution From Raw Image Bursts 从空间和时间上略有不同的视点捕获的多个低分辨率快照中重建高分辨率图像
10. Hierarchical Conditional Flow: A Unified Framework for Image Super-Resolution and Image Rescaling 图像超分辨率和缩放统一框架
11. Learning Frequency-Aware Dynamic Network for Efficient Super-Resolution
12. Dual-Camera Super-Resolution With Aligned Attention Modules 双摄像头超分辨率数据集，提出一种针对真实图像的自监督DA

### 图像绘制

1. CR-Fill: Generative Image Inpainting With Auxiliary Contextual Reconstruction 图像补全，联合训练辅助语境重建任务
2. Painting From Part 从部分的图像开始绘制并输出
3. Parallel Multi-Resolution Fusion Network for Image Inpainting
4. WaveFill: A Wavelet-Based Generation Network for Image Inpainting

### 图像恢复

1. Spatially-Adaptive Image Restoration Using Distortion-Guided Networks 利用失真定位信息并根据图像中困难区域动态调整计算
2. Unsupervised Non-Rigid Image Distortion Removal via Grid Deformation 无监督网络恢复潜在的无失真图像
3. Deep Reparametrization of Multi-Frame Super-Resolution and Denoising 多帧图像恢复

### 图像重建

1. High-Fidelity Pluralistic Image Completion With Transformers transformer进行外观先验重建
2. Focal Frequency Loss for Image Reconstruction and Synthesis 缩小频率的间隙可以进一步改善图像重建和合成质量
3. Semantic-Embedded Unsupervised Spectral Reconstruction From Single RGB Images in the Wild 单个RGB图像重建高光谱图像

### 图像质量评估

1. Learning Conditional Knowledge Distillation for Degraded-Reference Image Quality Assessment 针对原始质量图像不可用的情况提出的IQA方法
2. MUSIQ: Multi-Scale Image Quality Transformer 多尺度图像质量转换器处理不同大小和纵横比的图像

### 物体定位

1. Online Refinement of Low-Level Feature Based Activation Map for Weakly Supervised Object Localization 利用lowlevel特征定位，而不是主流的基于CAM的方法
2. Normalization Matters in Weakly Supervised Object Localization
3. Foreground Activation Maps for Weakly Supervised Object Localization 提出一种新的激活映射，联合优化弱监督物体定位中的两个任务，增强性能
4. Contrastive Attention Maps for Self-Supervised Co-Localization 引入了对比学习

### 密集预测

用预定义的类标记图像中的每个像素

1. FaPN: Feature-Aligned Pyramid Network for Dense Image Prediction 现有方法忽略了特征对齐的问题

### 图像转换

1. Bridging the Gap Between Label- and Reference-Based Synthesis in Multi-Attribute Image-to-Image Translation
2. Scaling-Up Disentanglement for Image Translation 图像从source域转换到目标域
3. GAN Inversion for Out-of-Range Images With Geometric Transformations
4. Instance-Wise Hard Negative Example Generation for Contrastive Learning in Unpaired Image-to-Image Translation 又说到了对比学习中负面例子的影响，训练一个生成器在线生成负面示例
5. Harnessing the Conditioning Sensorium for Improved Image Translation
6. Unaligned Image-to-Image Translation by Learning to Reweight
7. Frequency Domain Image Translation: More Photo-Realistic, Better Identity-Preserving 图像分解为低频和高频分量

### 图像匹配

1. P2-Net: Joint Description and Detection of Local Features for Pixel and Point Matching 像素关键点与3D关键点匹配
2. Multi-Scale Matching Networks for Semantic Correspondence 在一对语义相似的图像间寻找像素上的对应关系
3. Viewpoint Invariant Dense Matching for Visual Geolocalization 密集局部特征匹配对于视点偏移的鲁棒性比较差，在特征提取过程中嵌入视点移动不变性进行改进
4. Deep Matching Prior: Test-Time Optimization for Dense Correspondence 在一对输入图像上优化未经训练的匹配网络即可捕获特定于图像的先验
5. Learning To Match Features With Seeded Graph Matching Network 稀疏结构的图神经网络，减少冗余连接和学习紧凑表示
6. Warp Consistency for Unsupervised Learning of Dense Correspondences 之前无监督的方法难以应对大量的外观变化，依赖合成数据难以推广到真实数据，提出新的无监督学习目标
7. Matching in the Dark: A Dataset for Matching Image Pairs of Low-Light Scenes 弱光环境下的图像匹配数据集
8. COTR: Correspondence Transformer for Matching Across Images
9. OSCAR-Net: Object-Centric Scene Graph Attention for Image Attribution

### 图像压缩

1. Variable-Rate Deep Image Compression Through Spatially-Adaptive Feature Transform
2. Neural Image Compression via Attentional Multi-Scale Back Projection and Frequency Decomposition

### 成像

1. Self-Supervised Neural Networks for Spectral Snapshot Compressive Imaging 自监督光谱快照压缩成像，基本不需要标签数据
2. Accelerating Atmospheric Turbulence Simulation via Learned Phase-to-Space Transform 模拟大气湍流成像
3. Time-Multiplexed Coded Aperture Imaging: Learned Coded Aperture and Pixel Exposures for Compressive Imaging Systems 编码孔径压缩成像，成像一般都是硬件和算法协同设计
4. Dense Deep Unfolding Network With 3D-CNN Prior for Snapshot Compressive Imaging
5. Single-Shot Hyperspectral-Depth Imaging With Learned Diffractive Optics
6. Universal and Flexible Optical Aberration Correction Using Deep-Prior Based Deconvolution
7. Virtual Light Transport Matrices for Non-Line-of-Sight Imaging
8. Extreme-Quality Computational Imaging via Degradation Framework

### 图像全景拼接

1. Minimal Solutions for Panoramic Stitching Given Gravity Prior

### 相机定位、姿态回归

1. Viewing Graph Solvability via Cycle Consistency 利用循环一致性减少未知参数
2. Continual Learning for Image-Based Camera Localization 持续学习环境下的定位，缓冲区解决遗忘问题
3. PoGO-Net: Pose Graph Optimization With Graph Neural Networks
4. Learning Multi-Scene Absolute Pose Regression With Transformers 绝对相机姿态回归，仅使用图像来估计相机的位置和方向
5. On the Limits of Pseudo Ground Truth in Visual Camera Re-Localisation 质疑重新定位文献中的共同认知，即基于学习的场景坐标回归优于经典的基于特征的方法
6. Single View Physical Distance Estimation Using Human Pose RGB图像或者视频估计相机配置、距离
7. ViewNet: Unsupervised Viewpoint Estimation From Conditional Generation

### 图像异常检测

1. Image Manipulation Detection by Multi-View Multi-Scale Supervision 图像篡改检测(输出哪些区域被篡改) 多视角特征学习和多尺度监督
2. Reality Transform Adversarial Generators for Image Splicing Forgery Detection and Localization 对抗性训练生成和检测伪造图像
3. A Hierarchical Transformation-Discriminating Generative Model for Few Shot Anomaly Detection 分层生成模型捕获训练图像的patch分布
4. Learning Self-Consistency for Deepfake Detection 利用伪造图像中源特征的不一致性来检测假图像
5. Divide-and-Assemble: Learning Block-Wise Memory for Unsupervised Anomaly Detection 基于重建的图像无监督异常检测，将重建解释为
6. DRAEM - A Discriminatively Trained Reconstruction Embedding for Surface Anomaly Detection
7. Learning Unsupervised Metaformer for Anomaly Detection

### 视觉文本识别

1. Towards the Unseen: Iterative Text Recognition by Distilling From Errors 新框架解决难以识别未见过或稀少的字符序列
2. Joint Visual Semantic Reasoning: Multi-Stage Decoder for Text Recognition 带有视觉语义推理的文本识别
3. Text Is Text, No Matter What: Unifying Text Recognition Using Knowledge Distillation
4. Adaptive Boundary Proposal Network for Arbitrary Shape Text Detection
5. From Two to One: A New Scene Text Recognizer With Visual Language Modeling Network

### 视觉问答

从图像中回答开放式问题

1. Calibrating Concepts and Operations: Towards Symbolic Reasoning on Real Images 神经符号方法在真实图像上的改进
2. Weakly Supervised Relative Spatial Reasoning for Visual Question Answering
3. Linguistically Routing Capsule Network for Out-of-Distribution Visual Question Answering
4. Beyond Question-Based Biases: Assessing Multimodal Shortcut Learning in Visual Question Answering 提出一种评估VQA模型的方法

### 图像外推

1. SemIE: Semantically-Aware Image Extrapolation 图像外推，将输入图像扩展到边界之外，本文基于上下文在扩展区域添加新对象
2. Towards Complete Scene and Regular Shape for Distortion Rectification by Curve-Aware Extrapolation 广角镜头图片外推

### 光流估计

1. Learning To Estimate Hidden Motions With Global Motion Aggregation 只出现一次的点的光流估计，图像自相似性建模
2. Separable Flow: Learning Motion Cost Volumes for Optical Flow Estimation 光流估计
3. High-Resolution Optical Flow From 1D Attention and Correlation
4. GyroFlow: Gyroscope-Guided Unsupervised Optical Flow Learning 无监督，解决基本光流假设在不同场景下被打破的问题，提出一个新的数据集

### 视觉定位

1. Pose Correction for Highly Accurate Visual Localization in Large-Scale Indoor Spaces
2. CrowdDriven: A New Challenging Dataset for Outdoor Visual Localization 视觉定位：已知场景中估计给定图像的位置和方向

### 视觉关系

1. Few-Shot Visual Relationship Co-Localization 视觉关系共定位VRC
2. Exploring Long Tail Visual Relationship Recognition With Large Vocabulary 长尾视觉关系识别
3. Learning of Visual Relations: The Devil Is in the Tails

### 图像增强

1. Deep Symmetric Network for Underexposed Image Enhancement With Recurrent Attentional Learning 欠曝光图像增强问题视为原图与其增强版本之间的特征转换
2. Representative Color Transform for Image Enhancement
3. Adaptive Unfolding Total Variation Network for Low-Light Image Enhancement 恢复弱光的图像，在不放大噪声的情况下增加图像亮度
4. Real-Time Image Enhancer via Learnable Spatial-Aware 3D Lookup Tables 图像增强算法加速
5. STAR: A Structure-Aware Lightweight Transformer for Real-Time Image Enhancement 实时图像增强

### 其他

1. Learning To Resize Images for Computer Vision Tasks 研究图像大小的影响，训练一个图像尺寸调整器
2. Latent Transformations via NeuralODEs for GAN-Based Image Editing
3. Motion Basis Learning for Unsupervised Deep Homography Estimation With Subspace Projection 无监督深度单应估计，单应矩阵用于图像配准
4. LocalTrans: A Multiscale Local Transformer Network for Cross-Resolution Homography Estimation 交叉分辨率单应估计看作多模态问题
5. TransView: Inside, Outside, and Across the Cropping View Boundaries 图像裁剪推荐，CNN对于元素间关系的建模很弱，使用了transformer
6. AINet: Association Implantation for Superpixel Segmentation 图像超像素分割，提出新的模块，更好建模像素网格的上下文关系
7. On Equivariant and Invariant Learning of Object Landmark Representations 无监督地标表征学习，通过对比学习揭示学习不变量的本质
8. IICNet: A Generic Framework for Reversible Image Conversion 可逆图像转化，建立特定视觉内容和嵌入图像之间的可逆转换，该框架通用性能好
9. From Culture to Clothing: Discovering the World Events Behind a Century of Fashion Images 识别影响人类穿着的特定文化因素
10. DTMNet: A Discrete Tchebichef Moments-Based Deep Neural Network for Multi-Focus Image Fusion 多聚焦图像融合（组合两个或多个不同焦距拍摄的图像来获得全聚焦图像）
11. End-to-End Piece-Wise Unwarping of Document Images 去除纸张物理变形，便于OCR等任务
12. Embedding Novel Views in a Single JPEG Image
13. Attentional Pyramid Pooling of Salient Visual Residuals for Place Recognition 视觉位置识别（VPR）
14. Field of Junctions: Extracting Boundary Structure at Low SNR 寻找图像中的边界元素，对高噪声有鲁棒性
15. Hand Image Understanding via Deep Multi-Task Learning 手部图像理解新框架，级联多任务学习
16. Deep Halftoning With Reversible Binary Pattern 半色调？算法，将彩色图像转换并可以完全恢复到原始版本
17. Progressive Correspondence Pruning by Consensus Learning 像素间对应关系剪枝，剪错误，保留正确的
18. Multispectral Illumination Estimation Using Deep Unrolling Network 多光谱图像中的光照光谱估计问题
19. MDETR - Modulated Detection for End-to-End Multi-Modal Understanding 改进依靠预训练的detector的多模态推理系统，设计了一种端到端到框架
20. Calibrated and Partially Calibrated Semi-Generalized Homographies 广义相机的单应估计
21. Detecting Persuasive Atypicality by Modeling Contextual Compatibility 判断一张图片是否非典型，视觉广告数据集
22. Seasonal Contrast: Unsupervised Pre-Training From Uncurated Remote Sensing Data imageNet有领域差异，用于遥感相关任务的预训练数据集
23. Learning RAW-to-sRGB Mappings With Inaccurately Aligned Supervision 将输入的原始图像模仿成另一个相机捕获的sRGB图像，考虑到了图像对齐和颜色不一致问题带来的影响
24. SSH: A Self-Supervised Framework for Image Harmonization 图像协调，匹配前景和背景图像，从表示融合的角度重新描述图像协调问题
25. Planar Surface Reconstruction From Sparse Views 稀疏视角（比如几张视角重叠得不多的照片）进行平面重建
26. Cascade Image Matting With Deformable Graph Refinement 人体图像抠图
27. SUNet: Symmetric Undistortion Network for Rolling Shutter Correction 图像的相机RS效应消除
28. PixelPyramids: Exact Inference Models From Lossless Image Pyramids 自然图像密度估计任务
29. SketchAA: Abstract Representation for Abstract Sketches 针对素描设计的表示
30. Dynamic Cross Feature Fusion for Remote Sensing Pansharpening 遥感图像泛锐化
31. Deep Edge-Aware Interactive Colorization Against Color-Bleeding Effects 图像自动着色
32. ELSD: Efficient Line Segment Detector and Descriptor 为SLAM、图像匹配等更高层次的任务实时提供必要线条特征
33. Estimating and Exploiting the Aleatoric Uncertainty in Surface Normal Estimation 从单张图像估计曲面法线
34. The Functional Correspondence Problem 寻找图像对应关系的数据集
35. MT-ORL: Multi-Task Occlusion Relationship Learning 图像中检索对象之间的遮挡关系
36. Cross-Camera Convolutional Color Constancy 从训练期间没出现的相机拍摄的照片中估计场景的光源颜色
37. Who's Waldo? Linking People Across Text and Images 图像中观察到的人和文本中提到的人的关系
38. Unsupervised Layered Image Decomposition Into Object Prototypes 无监督学习框架，用于将图像分解为原型图像的显式变换
39. Do Image Classifiers Generalize Across Time? 图像分类器对时间扰动的鲁棒性数据集
40. Hierarchical Disentangled Representation Learning for Outdoor Illumination Estimation and Editing 室外光照预测
41. Image Harmonization With Transformer 图像协调，使合成图像看起来更逼真
42. Learning To Discover Reflection Symmetry via Polar Matching Convolution 反射对称性检测的任务是通过检测图像的对称轴来发现图像中的反射模式
43. SketchLattice: Latticed Representation for Sketch Manipulation 草图表示设计
44. Multi-Level Curriculum for Training a Distortion-Aware Barrel Distortion Rectification Model 消除广角镜头拍摄的畸变图像中的径向畸变

## 人脸相关

1. Understanding and Mitigating Annotation Bias in Facial Expression Recognition 关注标签偏差、注释偏差的面部表情识别框架
2. VariTex: Variational Neural Face Textures 提供了对语义上有意义的参数的独立控制的人脸图像生成
3. Learn To Cluster Faces via Pairwise Classification 从pairwise角度探讨人脸聚类
4. Learning Facial Representations From the Cycle-Consistency of Face 从人脸图像中分离和提取特征
5. Towards Face Encryption by Generating Adversarial Identity Masks 对人脸进行加密，但不影响其他任务的性能
6. Super-Resolving Cross-Domain Face Miniatures by Peeking at One-Shot Exemplar 人脸超分辨率
7. A Latent Transformer for Disentangled Face Editing in Images and Videos 面部图像编辑
8. Adaptive Label Noise Cleaning With Meta-Supervision for Deep Face Recognition 学习待清洗数据分布，自适应调整，先从标记良好的噪声数据中学习清洗知识，再通过元监督转移到目标数据
9. Disentangled Representation for Age-Invariant Face Recognition: A Mutual Information Minimization Perspective 解纠缠表征学习，多任务学习框架，最小化来自同一个人的人脸图像的身份分量和年龄分量之间的互信息
10. Personalized and Invertible Face De-Identification by Disentangled Identity Information Manipulation 人脸个性化可逆去识别和恢复
11. Retrieve in Style: Unsupervised Facial Feature Transfer and Retrieval 在图像上传输和检索特征
12. Talk-To-Edit: Fine-Grained Facial Editing via Dialog 人脸编辑
13. PASS: Protected Attribute Suppression System for Mitigating Bias in Face Recognition
14. Towards High Fidelity Monocular Face Reconstruction With Rich Reflectance Using Self-Supervised Learning and Ray Tracing 人脸重建
15. OpenForensics: Large-Scale Challenging Dataset for Multi-Face Forgery Detection and Segmentation In-the-Wild 野外多人脸伪造检测和分割的数据集
16. Re-Aging GAN: Toward Personalized Face Age Transformation 人脸老化
17. SynFace: Face Recognition With Synthetic Data

## 风格迁移

1. Learning To Stylize Novel Views 3D场景的风格迁移
2. Diverse Image Style Transfer via Invertible Cross-Space Mapping 多样性图像样式转换框架
3. Domain-Aware Universal Style Transfer 现有工作定义的任意风格有结构上的局限性，针对此进行改进
4. AdaAttN: Revisit Attention Mechanism in Arbitrary Neural Style Transfer
5. Manifold Alignment for Semantically Aligned Style Transfer

## 人群计数

1. Exploiting Sample Correlation for Crowd Counting With Multi-Expert Network 如何更高效地培养和选择专家网络
2. Spatial Uncertainty-Aware Semi-Supervised Crowd Counting 半监督比较受关注，因为全监督数据集太难做
3. Variational Attention: Propagating Domain-Specific Knowledge for Multi-Domain Learning in Crowd Counting 多个不同数据集训练，领域知识传播
4. Crowd Counting With Partial Annotations in an Image 新setting，部分注释训练，充分利用不同视角不同场景获得的数据
5. Towards a Universal Model for Cross-Dataset Crowd Counting 模型对规模变化有灾难性敏感，尺度对齐
6. Rethinking Counting and Localization in Crowds: A Purely Point-Based Framework 联合人群计数和个体定位框架

## 轨迹预测

1. Unlimited Neighborhood Interaction for Heterogeneous Trajectory Prediction
2. DenseTNT: End-to-End Trajectory Prediction From Dense Goal Sets 无需预先定义的锚的端到端模型
3. Where Are You Heading? Dynamic Trajectory Prediction With Expert Goal Examples
4. From Goals, Waypoints & Paths to Long Term Human Trajectory Forecasting 分解不确定性，两部分进行建模
5. LOKI: Long Term and Key Intentions for Trajectory Prediction 新的轨迹预测数据集
6. Three Steps to Multimodal Trajectory Prediction: Modality Clustering, Classification and Synthesis 多模态轨迹预测，新框架
7. Personalized Trajectory Prediction via Distribution Discrimination 分布判别预测个性化的运动模式
8. AgentFormer: Agent-Aware Transformers for Socio-Temporal Multi-Agent Forecasting
9. Spatial-Temporal Consistency Network for Low-Latency Trajectory Forecasting 更好地进行时间层次的建模
10. MG-GAN: A Multi-Generator Model Preventing Out-of-Distribution Samples in Pedestrian Trajectory Prediction

## 目标检测

1. FOVEA: Foveated Image Magnification for Autonomous Navigation 利用先验信息自动放大认为有小物体的区域
2. Query Adaptive Few-Shot Object Detection With Heterogeneous Graph Convolutional Networks few-shot目标检测 异构图卷积网络
3. Morphable Detector for Object Detection on Demand
4. Exploring Classification Equilibrium in Long-Tailed Object Detection 提出两个模块来平衡分类，关注长尾分布的问题
5. DetCo: Unsupervised Contrastive Learning for Object Detection 自监督对比学习
6. Domain-Invariant Disentangled Network for Generalizable Object Detection 目标检测泛化工作
7. Universal-Prototype Enhancing for Few-Shot Object Detection 利用不同对象类别中普遍存在的本质特征
8. Knowledge Mining and Transferring for Domain Adaptive Object Detection 新的框架用于DA
9. MFNet: Multi-Filter Directive Network for Weakly Supervised Salient Object Detection 生成的单个标签有偏见，提出多伪标签框架
10. **Robust Small Object Detection on the Water Surface Through Fusion of Camera and Millimeter Wave Radar 新的毫米波雷达点云表示格式与RGB图像特征融合**
11. Multitask AET With Orthogonal Tangent Regularity for Dark Object Detection
12. Seeking Similarities Over Differences: Similarity-Based Domain Alignment for Adaptive Object Detection 使检测器适应输入分布的变化，诱导源领和目标域的特征对齐
13. Deep Structured Instance Graph for Distilling Object Detectors 设计了一种图作为编码的知识形式从教师传递到学生
14. Wanderlust: Online Continual Object Detection in the Real World 提供一个新的在线连续目标检测benchmark
15. PreDet: Large-Scale Weakly Supervised Pre-Training for Detection 提出一种新的大规模预训练策略
16. Training Multi-Object Detector by Estimating Bounding Box Distribution for Input Image 将多目标检测转化为BBOX密度估计问题
17. FMODetect: Robust Detection of Fast Moving Objects 快速移动对象的鲁棒检测
18. Multi-Source Domain Adaptation for Object Detection 多源域DA
19. Mutual Supervision for Dense Object Detection 新的监督范式，相互监督
20. Self-Supervised Object Detection via Generative Image Synthesis 自监督目标检测，使用可控gan合成数据训练
21. SimROD: A Simple Adaptation Method for Robust Object Detection 无监督，客服域移位和伪标签噪声
22. Towards Real-World Prohibited Item Detection: A Large-Scale X-Ray Benchmark X射线违禁物品检测数据集
23. Rethinking Transformer-Based Set Prediction for Object Detection
24. Boosting Weakly Supervised Object Detection via Learning Bounding Box Adjusters 弱监督对象检测最近出现的topic 带良好注释的辅助数据集
25. CaT: Weakly Supervised Object Detection With Category Transfer
26. Dynamic DETR: End-to-End Object Detection With Dynamic Attention
27. CrossDet: Crossline Representation for Object Detection 不同于矩形锚框和一组点的表示方法
28. Robust Object Detection via Instance-Level Temporal Cycle Confusion
29. Semantically Coherent Out-of-Distribution Detection 分布外检测新benchmark，关注语义一致性
30. End-to-End Semi-Supervised Object Detection With Soft Teacher 半监督目标检测，教师是学生在不同步骤的聚合，提出了训练中两种简单而有效的技术
31. MosaicOS: A Simple and Effective Use of Object-Centric Images for Long-Tailed Object Detection
32. GraphFPN: Graph Feature Pyramid Network for Object Detection 提出了一种图特征金字塔网络，它能够使其拓扑结构适应不同的内在图像结构，并支持所有尺度上的同时特征交互
33. Reconcile Prediction Consistency for Balanced Object Detection
34. DeFRCN: Decoupled Faster R-CNN for Few-Shot Object Detection

## 边缘检测

1. Pixel Difference Networks for Efficient Edge Detection 简单、轻量级但有效
2. RINDNet: Edge Detection for Discontinuity in Reflectance, Illumination, Normal and Depth 联合检测四种类型的边缘

## 边缘计算

1. Real-Time Video Inference on Edge Devices via Adaptive Model Streaming 远程服务器在线知识蒸馏从大型模型中提升边缘设备运行的小型模型的性能

## 3D重建

1. THUNDR: Transformer-Based 3D Human Reconstruction With Markers 给定3D单目图像进行重建
2. PR-RRN: Pairwise-Regularized Residual-Recursive Networks for Non-Rigid Structure-From-Motion RGB图像重建3D形状和相机
3. Probabilistic Modeling for Human Mesh Recovery 基于2D信息的3D人体重建
4. Learning To Regress Bodies From Images Using Differentiable Semantic Rendering 引入服装语义信息，单目图像回归三维人体形状和姿态
5. VolumeFusion: Deep Depth Fusion for 3D Scene Reconstruction 利用DNN复刻传统方法，提高可解释性和准确性
6. In-the-Wild Single Camera 3D Reconstruction Through Moving Water Surfaces 根据水上的单目相机重建水下环境
7. Vis2Mesh: Efficient Mesh Reconstruction From Unstructured Point Clouds of Large Scenes With Learned Virtual View Visibility 非结构化点云重建mesh，在训练上进行改进
8. 3D Human Texture Estimation From a Single Image With Transformers 单目3D人体纹理估计
9. I2UV-HandNet: Image-to-UV Prediction Network for Accurate and High-Fidelity 3D Hand Mesh Modeling 从RGB图像恢复3D手部mesh
10. Geometric Granularity Aware Pixel-To-Mesh 几何粒度感知
11. Learning Efficient Photometric Feature Transform for Multi-View Stereo 学习光度信息包含的低层特征来增强重建表现
12. Self-Supervised Transfer Learning for Hand Mesh Recovery From Binocular Images 自监督迁移学习，双目图像手部重建
13. Learning Anchored Unsigned Distance Functions With Gradient Direction Alignment for Single-View Garment Reconstruction 无符号距离？单视图服装重建
14. Skeleton2Mesh: Kinematics Prior Injected Unsupervised Human Mesh Recovery 改进姿势不匹配和形状模糊两个问题，引入运动学先验
15. PlaneTR: Structure-Guided Transformers for 3D Plane Recovery seq2seq联合利用上下文信息和几何结构
16. TravelNet: Self-Supervised Physically Plausible Hand Motion Learning From Monocular Color Images 单目RGB图像重建物理上合理的手部运动
17. Gravity-Aware Monocular 3D Human-Object Reconstruction 单目RGB视频无关节标记3D人体运动捕捉
18. UltraPose: Synthesizing Dense Pose With 1 Billion Points by Human-Body Decoupling 3D Model 图像中恢复密集的人体姿势，建立图像到3D表面的对应关系
19. Just a Few Points Are All You Need for Multi-View Stereo: A Novel Semi-Supervised Learning Method for Multi-View Stereo 从多视点图像中重建密集的三维几何体，半监督，放松监督，缓解有限MVS数据的阻碍
20. GTT-Net: Learned Generalized Trajectory Triangulation
21. Deep Permutation Equivariant Structure From Motion
22. 3DeepCT: Learning Volumetric Scattering Tomography of Clouds
23. SA-ConvONet: Sign-Agnostic Optimization of Convolutional Occupancy Networks 3D点云曲面重建
24. AA-RMVSNet: Adaptive Aggregation Recurrent Multi-View Stereo Network 多视角3D重建，针对复杂场景中遮挡变化的困难
25. Self-Supervised 3D Face Reconstruction via Conditional Estimation 通过视频的自监督训练，从2D单目图像学习3D人脸参数
26. Multiresolution Deep Implicit Functions for 3D Shape Representation
27. Towers of Babel: Combining Images, Language, and 3D Geometry for Learning Multimodal Vision 新的数据集，涉及图像、文本和3D几何的多模态推理
28. Discovering 3D Parts From Image Collections 自监督单视图图像推理3D形状
29. Gaussian Fusion: Accurate 3D Reconstruction via Geometry-Guided Displacement Interpolation
30. Digging Into Uncertainty in Self-Supervised Multi-View Stereo
31. Uncertainty-Aware Human Mesh Recovery From Video by Learning Part-Based 3D Dynamics 同时预测3D姿势和3D运动学
32. DensePose 3D: Lifting Canonical Surface Maps of Articulated Objects to the Third Dimension 铰接物体单目3D重建
33. Deep Implicit Surface Point Prediction Networks
34. PatchMatch-RL: Deep MVS With Pixelwise Depth, Normal, and Visibility
35. Rational Polynomial Camera Model Warping for Deep Learning Based Satellite Multi-View Stereo Matching
36. ARCH++: Animation-Ready Clothed Human Reconstruction Revisited 基于之前框架的局限进行改进
37. RetrievalFuse: Neural 3D Scene Reconstruction With a Database
38. Generative Compositional Augmentations for Scene Graph Prediction 场景图形式的图像中推断对象及其关系，扰动真实场景图来合成新数据
39. PX-NET: Simple and Efficient Pixel-Wise Training of Photometric Stereo Networks 物体反射光获取精确3D重建
40. Dynamic Surface Function Networks for Clothed Human Bodies
41. A-SDF: Learning Disentangled Signed Distance Functions for Articulated Shape Representation
42. Learning Signed Distance Field for Multi-View Surface Reconstruction
43. Multi-View 3D Reconstruction With Transformers
44. A Confidence-Based Iterative Solver of Depths and Surface Normals for Deep Multi-View Stereo
45. Lightweight Multi-Person Total Motion Capture Using Sparse Multi-View Cameras 多视角相机多人运动捕捉输出人体模型
46. Monocular, One-Stage, Regression of Multiple 3D People
47. H3D-Net: Few-Shot High-Fidelity 3D Head Reconstruction
48. 3DIAS: 3D Shape Reconstruction With Implicit Algebraic Surfaces
49. Polarimetric Helmholtz Stereopsis
50. 3D Building Reconstruction From Monocular Remote Sensing Images

## ReID

1. Cross-Modality Person Re-Identification via Modality Confusion and Center Aggregation 跨模态ReID
2. Clothing Status Awareness for Long-Term Person Re-Identification 服装状态感知
3. TransReID: Transformer-Based Object Re-Identification 第一篇使用纯transformer的reid
4. Self-Supervised Geometric Features Discovery via Interpretable Attention for Vehicle Re-Identification and Beyond 车辆ReID，自监督表示学习促进几何特征的发现
5. Explainable Person Re-Identification With Attribute-Guided Metric Distillation 属性引导度量蒸馏，增加模型可解释性
6. Dense Interaction Learning for Video-Based Person Re-Identification 利用多尺度细粒度特征并构建它们之间的结构交互
7. Heterogeneous Relational Complement for Vehicle Re-Identification 构造鲁棒特征表示和提出相机敏感评价，学习cross-view表示
8. Online Pseudo Label Generation by Hierarchical Cluster Dynamics for Adaptive Person Re-Identification 改进生成伪标签的方法
9. Pyramid Spatial-Temporal Aggregation for Video-Based Person Re-Identification 分层时间特征融合
10. ICE: Inter-Instance Contrastive Encoding for Unsupervised Person Re-Identification 无监督，关注实例层面的对比损失
11. BV-Person: A Large-Scale Dataset for Bird-View Person Re-Identification
12. Learning To Know Where To See: A Visibility-Aware Approach for Occluded Person Re-Identification 对额外信息的稀疏和噪声更加鲁棒
13. Meta Pairwise Relationship Distillation for Unsupervised Person Re-Identification
14. Syncretic Modality Collaborative Learning for Visible Infrared Person Re-Identification
15. Video-Based Person Re-Identification With Spatial and Temporal Memory Networks 空间记忆存储了经常出现在视频帧中的空间干扰因素的特征，而时间记忆则保存了针对个人视频中典型时间模式进行优化的注意力
16. Visio-Temporal Attention for Multi-Camera Multi-Target Association 多目标多摄像机中的ReID

## 行人相关

### 行人行为预测

1. Bifold and Semantic Reasoning for Pedestrian Behavior Prediction 多任务学习框架，依赖多模态数据

### 行人检测

1. Stacked Homography Transformations for Multi-View Pedestrian Detection

### 行人属性识别

预测行人的多种属性作为行人图像的语义描述

1. Spatial and Semantic Consistency Regularizations for Pedestrian Attribute Recognition

## 3D目标检测

1. 4D-Net for Learned Multi-Modal Alignment 使用对应的多帧点云和RGB图像进行学习
2. Fog Simulation on Real LiDAR Point Clouds for 3D Object Detection in Adverse Weather 模拟大雾天气
3. An End-to-End Transformer Model for 3D Object Detection 易实现，据称适用于检测之外的3D任务
4. Unsupervised Domain Adaptive 3D Detection With Multi-Level Consistency 认为3D检测跨域gap是几何不匹配
5. Group-Free 3D Object Detection via Transformers 不采用人工制作的分组方案，使用transformer的注意力机制
6. Improving 3D Object Detection With Channel-Wise Transformer 关键点进行空间上下文建模
7. Geometry Uncertainty Projection Network for Monocular 3D Object Detection 不确定性推理加分层任务学习，减少几何杠杆负面影响
8. VENet: Voting Enhancement Network for 3D Object Detection 增强了投票功能
9. Multi-Echo LiDAR for 3D Object Detection 提出新的模块用于多回波雷达的数据学习
10. Voxel Transformer for 3D Object Detection
11. RangeDet: In Defense of Range View for LiDAR-Based 3D Object Detection 不是体素化或者BEV表示，距离视图表示，进行了改进
12. The Devil Is in the Task: Exploiting Reciprocal Appearance-Localization Features for Monocular 3D Object Detection 将单目3D检测转化为目标定位和外观感知子任务
13. Gated3D: Monocular 3D Object Detection From Temporal Illumination Cues 利用时间照明线索
14. FIERY: Future Instance Prediction in Bird's-Eye View From Surround Monocular Cameras 多相机3D检测
15. Are We Missing Confidence in Pseudo-LiDAR Methods for Monocular 3D Object Detection? 发现伪雷达方法的问题并改进
16. RandomRooms: Unsupervised Pre-Training From Synthetic Shapes and Randomized Layouts for 3D Object Detection 合成数据预训练方法

## 2D姿态估计

1. An Empirical Study of the Collapsing Problem in Semi-Supervised 2D Human Pose Estimation
2. End-to-End Detection and Pose Estimation of Two Interacting Hands 两只交互手的姿态估计
3. Keypoint Communities 将关键点建模为图
4. The Center of Attention: Center-Keypoint Grouping via Attention for Multi-Person Pose Estimation 应用transformer将关节分组到人中心
5. Online Knowledge Distillation for Efficient Pose Estimation 在线知识蒸馏架构，提高效率
6. Space-Time-Separable Graph Convolutional Network for Pose Forecasting
7. Multi-Instance Pose Networks: Rethinking Top-Down Pose Estimation
8. Human Pose Regression With Residual Log-Likelihood Estimation
9. Removing the Bias of Integral Pose Regression

## 3D动作预测

1. Contextually Plausible and Diverse 3D Human Motion Prediction 3D人体动作预测，保持了多样性
2. MSR-GCN: Multi-Scale Residual Graph Convolution Networks for Human Motion Prediction
3. Generating Smooth Pose Sequences for Diverse Human Motion Prediction
4. Motion Prediction Using Trajectory Cues 关节轨迹空间中建模运动上下文

## 3D姿态估计

1. EventHands: Real-Time Neural 3D Hand Pose Estimation From an Event Stream 单目视频3D手部姿态估计
2. DECA: Deep Viewpoint-Equivariant Human Pose Estimation Using Capsule Autoencoders
3. Shape-Aware Multi-Person Pose Estimation From Multi-View Images 多人多视角3D姿态估计
4. 3D Human Pose Estimation With Spatial and Temporal Transformers 对每一帧内的人体关节关系和帧间的时间相关性进行建模
5. SemiHand: Semi-Supervised Hand Pose Estimation With Consistency 半监督手部姿态估计
6. PARE: Part Attention Regressor for 3D Human Body Estimation 引入软注意力机制解决遮挡对于3D人体姿态和形状估计造成的影响
7. Learning Causal Representation for Training Cross-Domain Pose Estimator via Generative Interventions 因果表征学习
8. Learning Skeletal Graph Neural Networks for Hard 3D Pose Estimation
9. Skeleton2Mesh: Kinematics Prior Injected Unsupervised Human Mesh Recovery 3D
10. Modulated Graph Convolutional Network for 3D Human Pose Estimation 根据应用图卷积网络方法的缺点进行改进
11. Revitalizing Optimization for 3D Human Pose and Shape Estimation: A Sparse Constrained Formulation
12. Unsupervised 3D Pose Estimation for Hierarchical Dance Video Recognition
13. Hierarchical Kinematic Probability Distributions for 3D Human Shape and Pose Estimation From Images in the Wild
14. EM-POSE: 3D Human Pose Estimation From Sparse Electromagnetic Trackers
15. Towards Alleviating the Modeling Ambiguity of Unsupervised Monocular 3D Human Pose Estimation
16. Estimating Egocentric 3D Human Pose in Global Space
17. Camera Distortion-Aware 3D Human Pose Estimation in Video With Optimization-Based Meta-Learning
18. Self-Supervised 3D Hand Pose Estimation From Monocular RGB via Contrastive Learning
19. Egocentric Pose Estimation From Human Vision Span 自我为中心的视角估计相机佩戴者的身体姿势
20. GarmentNets: Category-Level Pose Estimation for Garments via Canonical Space Shape Completion
21. Holistic Pose Graph: Modeling Geometric Structure Among Objects in a Scene Using Graph Inference for 3D Object Prediction
22. HandFoldingNet: A 3D Hand Pose Estimation Network Using Multiscale-Feature Guided Folding of a 2D Hand Skeleton

## 6D姿态估计

1. SGPA: Structure-Guided Prior Adaptation for Category-Level 6D Object Pose Estimation
2. RePOSE: Fast 6D Object Pose Refinement via Deep Texture Rendering 改进方法加快特征提取速度和精度
3. DualPoseNet: Category-Level 6D Object Pose and Size Estimation Using Dual Pose Network With Refined Learning of Pose Consistency
4. SO-Pose: Exploiting Self-Occlusion for Direct 6D Pose Estimation 单目图像直接回归
5. PR-GCN: A Deep Graph Convolutional Network With Point Refinement for 6D Pose Estimation
6. Probabilistic Monocular 3D Human Pose Estimation With Normalizing Flows

## Transformer

1. 3DVG-Transformer: Relation Modeling for Visual Grounding on Point Clouds （visual grounding，输入image和对应的物体描述，输出物体box）
2. Describing and Localizing Multiple Changes With Transformers
3. Pyramid Vision Transformer: A Versatile Backbone for Dense Prediction Without Convolutions transformer backbone研究
4. Multi-Scale Vision Longformer: A New Vision Transformer for High-Resolution Image Encoding 利用transformer编码高分辨率图像
5. LeViT: A Vision Transformer in ConvNet's Clothing for Faster Inference 图像分类
6. Understanding Robustness of Transformers for Image Classification 理解ViT在图像分类下的鲁棒性
7. FuseFormer: Fusing Fine-Grained Information in Transformers for Video Inpainting
8. UniT: Multimodal Multitask Learning With a Unified Transformer 使用transformer的多任务学习
9. VidTr: Video Transformer Without Convolutions 适用于视频的transformer改进
10. PnP-DETR: Towards Efficient Visual Analysis With Transformers 减少冗余计算，图像特征映射为前景对象特征和少量的背景上下文特征
11. Visformer: The Vision-Friendly Transformer 逐步将基于transformer的模型变成基于卷积的模型
12. Cloud Transformers: A Universal Approach to Point Cloud Processing Tasks 根据点云特点做出了改进的transformer
13. Going Deeper With Image Transformers 更深的改进ViT
14. Multiscale Vision Transformers 多尺度特征层次的基本思想
15. Conditional DETR for Fast Training Convergence 加快训练速度
16. Rethinking Spatial Dimensions of Vision Transformers 研究空间维度体系架构对于ViT的影响
17. An Empirical Study of Training Self-Supervised Vision Transformers 自监督训练方法研究
18. On the Robustness of Vision Transformers to Adversarial Examples Vit对对抗攻击的鲁棒性
19. Tokens-to-Token ViT: Training Vision Transformers From Scratch on ImageNet
20. Rethinking and Improving Relative Position Encoding for Vision Transformer
21. Visual Saliency Transformer
22. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows
23. StyleFormer: Real-Time Arbitrary Style Transfer via Parametric Style Composition
24. Vision Transformers for Dense Prediction
25. CrackFormer: Transformer Network for Fine-Grained Crack Detection
26. Incorporating Convolution Designs Into Visual Transformers
27. Point Transformer
28. CvT: Introducing Convolutions to Vision Transformers 卷积特性引入，且保持原来的优点
29. ViViT: A Video Vision Transformer
30. Vision Transformer With Progressive Sampling 采样策略定位有区分性的区域
31. Fast Convergence of DETR With Spatially Modulated Co-Attention 改进注意力机制，加快收敛速度
32. Conformer: Local Features Coupling Global Representations for Visual Recognition 卷积和注意力结合，最大程度保留局部特征和全局表示

## NeRF、深度学习图形学结合

1. PlenOctrees for Real-Time Rendering of Neural Radiance Fields 利用PlenOctress改进NeRF
2. Neural Radiance Flow for 4D View Synthesis and Video Processing 一组RGB生成4D动态场景的时空表征
3. Unconstrained Scene Generation With Locally Conditioned Radiance Fields 将一个相对大的空间分解成多个局部的辐射场
4. Continual Neural Mapping: Learning an Implicit Scene Representation From Sequential Observations 连续观测中学习隐式场景表示
5. Animatable Neural Radiance Fields for Modeling Dynamic Human Bodies 多视角视频重建可动画人体模型，引入神经混合权重场
6. GNeRF: GAN-Based Neural Radiance Field Without Posed Camera 减少对精确相机姿态的依赖
7. BARF: Bundle-Adjusting Neural Radiance Fields 训练不完美相机姿态的Nerf
8. FastNeRF: High-Fidelity Neural Rendering at 200FPS 加速Nerf
9. Dynamic View Synthesis From Dynamic Monocular Video 动态场景单目视频训一个静态和一个动态nerf，以无监督方式混合结果
10. NeRD: Neural Reflectance Decomposition From Image Collections
11. Putting NeRF on a Diet: Semantically Consistent Few-Shot View Synthesis
12. Editing Conditional Radiance Fields 粗糙的2D用户涂鸦传播到3D，以修改局部区域的颜色或形状
13. GRF: Learning a General Radiance Field for 3D Representation and Rendering
14. Neural Articulated Radiance Field 学习关节对象的姿态可控表示
15. Learning Object-Compositional Neural Radiance Field for Editable Scene Rendering
16. KiloNeRF: Speeding Up Neural Radiance Fields With Thousands of Tiny MLPs 加速nerf
17. CodeNeRF: Disentangled Neural Radiance Fields for Object Categories 分离形状和纹理
18. Baking Neural Radiance Fields for Real-Time View Synthesis

## 实例分割

1. Instances As Queries 将感兴趣的目标视为可学习的查询
2. FASA: Feature Augmentation and Sampling Adaptation for Long-Tailed Instance Segmentation 增加特征空间解决数据稀疏
3. Prior to Segment: Foreground Cues for Weakly Annotated Classes in Partially Supervised Instance Segmentation 引入新的先验，弱监督weak box标签
4. DiscoBox: Weakly Supervised Instance Segmentation and Semantic Correspondence From Box Supervision 新框架，bbox标签联合学习实例分割和语义对应
5. The Surprising Impact of Mask-Head Architecture on Novel Class Segmentation 减少注释成本，所有类别使用bbox，子类使用mask
6. SOTR: Segmenting Objects With Transformers
7. A Weakly Supervised Amodal Segmenter With Boundary Uncertainty Estimation 目标是分割可见和遮挡（amodal）对象部分，而训练只提供可见（modal）部分的GT分割

## 3D实例分割

1. Instance Segmentation in 3D Scenes Using Semantic Superpoint Tree Networks 根据以往pipeline的缺点进行改进

## 语义分割

1. Entropy Maximization and Meta Classification for Out-of-Distribution Detection in Semantic Segmentation 针对OoD问题提出了一个两步程序来改进现有的方法
2. Panoptic Segmentation of Satellite Image Time Series With Convolutional Temporal Attention Networks 全景卫星时序图像分割，提供了数据集和一个端到端一阶段的模块
3. Generalize Then Adapt: Source-Free Domain Adaptive Semantic Segmentation 域泛化+目标适应语义分割
4. Region-Aware Contrastive Learning for Semantic Segmentation 对比学习语义分割
5. ACDC: The Adverse Conditions Dataset With Correspondences for Semantic Driving Scene Understanding 建立了一个新的数据集，不利视觉条件下的分割
6. Domain Adaptive Semantic Segmentation With Self-Supervised Depth Estimation 深度估计辅助任务的语义分割
7. Scribble-Supervised Semantic Segmentation by Uncertainty Reduction on Neural Representation and Self-Supervision on Neural Eigenspace scribble监督（在图片上的小部分区域进行标注），第一次碰到这个概念
8. Robust Trust Region for Weakly Supervised Segmentation
9. Exploring Cross-Image Pixel Contrast for Semantic Segmentation 关注以往忽略的不同图像像素间丰富的语义关系
10. Leveraging Auxiliary Tasks With Affinity Learning for Weakly Supervised Semantic Segmentation 显著性检测和多标签分类做辅助任务
11. Weakly Supervised Segmentation of Small Buildings With Point Labels
12. Self-Regulation for Semantic Segmentation 根据语义分割失败的情况进行改进，设计新的loss
13. Unsupervised Segmentation Incorporating Shape Prior via Generative Adversarial Networks 对于光照变化和遮挡的鲁棒性和性能增强
14. Scaling Up Instance Annotation via Label Propagation 研究自动标注分割mask
15. Re-Distributing Biased Pseudo Labels for Semi-Supervised Semantic Segmentation: A Baseline Investigation 针对长尾分布，提出无偏伪标签生成和渐进数据扩充与标记策略
16. Uncertainty-Aware Pseudo Label Refinery for Domain Adaptive Semantic Segmentation 不确定性感知伪标签精炼，解决有噪音的伪标签对于域迁移的影响
17. Mining Contextual Information Beyond Image for Semantic Segmentation 挖掘图像之外的上下文信息增强像素表示
18. Contrastive Learning for Label Efficient Semantic Segmentation 直接使用交叉熵容易过拟合少量的标记，先使用对比loss后再用交叉熵
19. BiMaL: Bijective Maximum Likelihood Approach to Domain Adaptation in Semantic Scene Segmentation 新的loss，无监督衡量模型在新目标域上的效率
20. BAPA-Net: Boundary Adaptation and Prototype Alignment for Cross-Domain Semantic Segmentation 正确处理跨域语义分割的对象边界
21. ECS-Net: Improving Weakly Supervised Semantic Segmentation by Using Connections Between Class Activation Maps 利用CAM间的关系提出一种新的弱监督语义分割方法
22. Simpler Is Better: Few-Shot Semantic Segmentation With Classifier Weight Transformer 简化元学习任务，将编解码器留给预训练
23. Unlocking the Potential of Ordinary Classifier: Class-Specific Adversarial Erasing Framework for Weakly Supervised Semantic Segmentation
24. Interaction via Bi-Directional Graph of Semantic Region Affinity for Scene Parsing 将图像的每个区域作为一个整体，捕捉结构拓扑和不同区域之间的亲和度
25. Few-Shot Semantic Segmentation With Cyclic Memory Network
26. Seminar Learning for Click-Level Weakly Supervised Semantic Segmentation 因click-level注释的监督和成本权衡较好，seminar学习是用来自不同网络的知识来弥补注释信息不足
27. SIGN: Spatial-Information Incorporated Generative Network for Generalized Zero-Shot Semantic Segmentation 基于知识蒸馏的自训练策略
28. Conditional Diffusion for Interactive Segmentation 带有人类交互的语义分割，本文是点击
29. Labels4Free: Unsupervised Segmentation Using StyleGAN 利用styleGAN生成数据的无监督分割
30. Complementary Patch for Weakly Supervised Semantic Segmentation CAM不足以作为伪masks
31. Hypercorrelation Squeeze for Few-Shot Segmentation
32. Exploiting a Joint Embedding Space for Generalized Zero-Shot Semantic Segmentation 大多数采用生成方法，将不可见类的视觉特征从对应语义特征中综合出来，但有局限性，提出了一种区别对待的方法，统一框架解决
33. Deep Metric Learning for Open World Semantic Segmentation 具有对比聚类的深度度量学习网络，开放集语义分割
34. Scaling Semantic Segmentation Beyond 1K Classes on a Single GPU
35. Vision-Language Transformer and Query Generation for Referring Segmentation referring分割，给出一个文本，找到相应地对象并生成mask，重新定义问题，在图像中找到query最受关注的区域
36. Pseudo-Mask Matters in Weakly-Supervised Semantic Segmentation 针对使用CAM生成伪mask和使用噪声伪mask监督产生的问题做出改进
37. An Elastica Geodesic Approach With Convexity Shape Prior 提出一个更复杂的问题
38. Human Detection and Segmentation via Multi-View Consensus 人体检测与分割
39. Joint Topology-Preserving and Feature-Refinement Network for Curvilinear Structure Segmentation 曲线结构分割，结合增强特征表示和保留拓扑结构
40. Enhanced Boundary Learning for Glass-Like Object Segmentation 类玻璃物体的分割
41. From Contexts to Locality: Ultra-High Resolution Image Segmentation via Locality-Aware Contextual Correlation 超高分辨率图像的分割
42. Local Temperature Scaling for Probability Calibration 多标签语义分割校正方法
43. **Prototypical Matching and Open Set Rejection for Zero-Shot Semantic Segmentation**
44. Self-Mutating Network for Domain Adaptive Segmentation in Aerial Images

## 3D语义分割

1. Perturbed Self-Distillation: Weakly Supervised Large-Scale Point Cloud Semantic Segmentation 弱监督点云语义分割，受自监督启发提出了一种新的扰动分支框架，同时集成上下文模块对标签点的affinity correlation（不懂）进行正则
2. GP-S3Net: Graph-Based Panoptic Sparse Semantic Segmentation Network 全景语义分割，在点云数据集上做的
3. Persistent Homology Based Graph Convolution Network for Fine-Grained 3D Shape Segmentation 提出新的神经网络学习拓普数据分析方法
4. VMNet: Voxel-Mesh Network for Geodesic-Aware 3D Semantic Segmentation
5. Weakly Supervised 3D Semantic Segmentation Using Cross-Image Consensus and Inter-Voxel Affinity Relations
6. PARTS: Unsupervised Segmentation With Slots, Attention and Independence Maximization

## 3D形状识别

1. Learning Canonical View Representation for 3D Shape Recognition With Arbitrary Views 从任意视角中识别三维形状
2. MVTN: Multi-View Transformation Network for 3D Shape Recognition 观察到目前的多视图方法中相机视角是启发式和固定设置的，建议学习这些视角

## 显著检测

1. Specificity-Preserving RGB-D Saliency Detection 提出一种特异性保持网络架构和两个新的模块达到SOTA
2. Salient Object Ranking With Position-Preserved Attention 显著目标排序
3. Summarize and Search: Learning Consensus-Aware Dynamic Convolution for Co-Saliency Detection 针对以往方法中的不足进行改进，共识感知动态卷积
4. Disentangled High Quality Salient Object Detection 针对以往方法的局限性，将框架分解为两个部分
5. RGB-D Saliency Detection via Cascaded Mutual Information Minimization 互信息最小化，又贡献一个数据集
6. Scene Context-Aware Salient Object Detection

## 点云

### 其他

1. Walk in the Cloud: Learning Curves for Point Clouds Shape Analysis 新的点云特征学习方式
2. Point-Set Distances for Learning Representations of 3D Point Clouds 点云表示学习
3. Spatio-Temporal Self-Supervised Representation Learning for 3D Point Clouds 点云时空表示自监督学习
4. Differentiable Convolution Search for Point Cloud Processing 根据点云特点设计的卷积网络
5. Temporally-Coherent Surface Reconstruction via Metric-Consistent Atlases 使用点云进行无监督时间相干表面重建
6. PICCOLO: Point Cloud-Centric Omnidirectional Localization 通过点云和全景图还原相机姿态
7. Superpoint Network for Point Cloud Oversegmentation 超点：在语义上对具有相似几何特征的点进行分组
8. PU-EVA: An Edge-Vector Based Approximation Solution for Flexible-Scale Point Cloud Upsampling 点云上采样，更密集规则地逼近目标物体
9. A Closer Look at Rotation-Invariant Deep Point Cloud Analysis 研究点云旋转不变性
10. Embed Me if You Can: A Geometric Perceptron 涉及点云的几何问题？
11. CSG-Stump: A Learning Friendly CSG-Like Representation for Interpretable Shape Parsing 从点云生成3D形状表示
12. Unsupervised Learning of Fine Structure Generation for 3D Point Clouds by 2D Projections Matching 没有3D监控的情况下生成3D点云，将3D点云学习转化为2D投影匹配
13. SGMNet: Learning Rotation-Invariant Point Cloud Representations via Sorted Gram Matrix 研究点云旋转不变表征
14. Vector Neurons: A General Framework for SO(3)-Equivariant Networks
15. SAT: 2D Semantics Assisted Training for 3D Visual Grounding
16. Learning Inner-Group Relations on Point Clouds
17. Self-Supervised Pretraining of 3D Features on Any Point-Cloud
18. CPFN: Cascaded Primitive Fitting Networks for High-Resolution Point Clouds
19. Robustness Certification for Point Cloud Models 点云模型鲁棒性验证

### 物体定位

1. Free-Form Description Guided 3D Visual Graph Network for Object Grounding in Point Cloud

### 点云去噪

1. Score-Based Point Cloud Denoising

## 点云分割

1. TempNet: Online Semantic Segmentation on Large-Scale Point Cloud Series
2. Guided Point Contrastive Learning for Semi-Supervised Point Cloud Semantic Segmentation 对比学习+半监督点云分割
3. PrimitiveNet: Primitive Instance Segmentation With Local Primitive Embedding Under Adversarial Metric 点云进行高分辨率原始实例分割，将全局分割转化为更简单的局部任务
4. HPNet: Deep Primitive Segmentation Using Hybrid Representations 点云表示的3D形状分割为原始patches
5. Learning With Noisy Labels for Robust Point Cloud Segmentation 点噪声自适应
6. DRINet: A Dual-Representation Iterative Learning Network for Point Cloud Segmentation 双表示迭代学习
7. ReDAL: Region-Based and Diversity-Aware Active Learning for Point Cloud Semantic Segmentation 减少注释负担，基于区域和多样性感知的主动学习
8. Unsupervised Point Cloud Object Co-Segmentation by Co-Contrastive Learning and Mutual Attention Sampling
9. RPVNet: A Deep and Efficient Range-Point-Voxel Fusion Network for LiDAR Point Cloud Segmentation

### 点云配准

1. Feature Interactive Representation for Point Cloud Registration
2. LSG-CPD: Coherent Point Drift With Local Surface Geometry for Point Cloud Registration 概率点云配准
3. Sampling Network Guided Cross-Entropy Method for Unsupervised Point Cloud Registration 建模为马尔科夫决策过程
4. A Robust Loss for Point Cloud Registration 提出一种新的度量方法
5. Deep Hough Voting for Robust Global Registration 利用6D Hough空间进行投票
6. HRegNet: A Hierarchical Network for Large-Scale Outdoor LiDAR Point Cloud Registration 分层网络，分层提取关键点
7. OMNet: Learning Overlapping Mask for Partial-to-Partial Point Cloud Registration
8. DeepPRO: Deep Partial Point Cloud Registration of Objects 无关键点，想象点云的形状后配准
9. (Just) A Spoonful of Refinements Helps the Registration Error Go Down
10. Provably Approximated Point Cloud Registration
11. Bootstrap Your Own Correspondences 自监督几何特征提取
12. Distinctiveness Oriented Positional Equilibrium for Point Cloud Registration GNN主干遭受着与高级特征的过度平滑和结构模糊相关的难以区分的特征问题，这是点云配准的一个关键瓶颈
13. PCAM: Product of Cross-Attention Matrices for Rigid Registration of Point Clouds

### 点云补全

1. Voxel-Based Network for Shape Completion by Leveraging Edge Generation
2. SnowflakeNet: Point Cloud Completion by Snowflake Point Deconvolution With Skip-Transformer
3. RFNet: Recurrent Forward Network for Dense Point Cloud Completion 新的递归前向网络，三个模块组成
4. PoinTr: Diverse Point Cloud Completion With Geometry-Aware Transformers 新setting，将补全定义为set2set翻译问题
5. ME-PCN: Point Completion Conditioned on Mask Emptiness 利用3D形状空间中的“空”点

### 目标追踪

1. Exploring Simple 3D Multi-Object Tracking for Autonomous Driving 可以去除启发式匹配步骤的端到端训练模型
2. Learning To Adversarially Blur Visual Object Tracking 针对运动模糊进行对抗攻击训练
3. Assignment-Space-Based Multi-Object Tracking and Segmentation
4. Learning Target Candidate Association To Keep Track of What Not To Track
5. A General Recurrent Tracking Framework Without Real Data 潜在的跟踪评分机制具有重要意义，新框架中通过长期信息对潜在轨迹进行评分，并且提出一种生成模拟数据的方法
6. MLVSNet: Multi-Level Voting Siamese Network for 3D Visual Tracking 受Hough投票的启发，多层次特征投票保留更多信息
7. Box-Aware Feature Enhancement for Single Object Tracking on Point Clouds 主流方法是基于目标模板和搜索区域之间的特征比较来跟踪目标，文章利用第一帧中的GT bbox增强目标对象特征，还提出一种表示方法用于点到框关系来描述对象
8. Learning To Track Objects From Unlabeled Videos
9. Saliency-Associated Object Tracking 研究之前策略的缺点，提出根据目标的显著部分进行追踪
10. HiFT: Hierarchical Feature Transformer for Aerial Tracking
11. Making Higher Order MOT Scalable: An Efficient Approximate Solver for Lifted Disjoint Paths
12. Transparent Object Tracking Benchmark 对透明物体的跟踪，新的数据集
13. DepthTrack: Unveiling the Power of RGBD Tracking RGBD跟踪数据集
14. Object Tracking by Jointly Exploiting Frame and Event Domain
15. Track Without Appearance: Learn Box and Tracklet Embedding With Local and Global Motion Patterns for Vehicle Tracking 车辆跟踪
16. Learning Spatio-Temporal Transformer for Visual Tracking
17. High-Performance Discriminative Tracking With Transformers 有效消除了手工设计的区分模型的需要
18. Continuous Copy-Paste for One-Stage Multi-Object Tracking and Segmentation 数据增强策略，增强一阶段方法的性能，超过了之前的二阶段方法

## Video相关

### Video动作分类

1. Composable Augmentation Encoding for Video Representation Learning 对比学习+自监督视频理解框架
2. Contrast and Order Representations for Video Self-Supervised Learning 自监督框架，实验做的是动作分类
3. HAA500: Human-Centric Atomic Action Dataset With Curated Videos 贡献一个新的以人为中心的细粒度动作数据集
4. Video Pose Distillation for Few-Shot, Fine-Grained Sports Action Recognition 利用姿势特征加强动作识别
5. Learning Self-Similarity in Space and Time As Generalized Motion for Video Action Recognition 基于时空自相似性的运动表示帮助学习视频中的运动动力学
6. Unsupervised Few-Shot Action Recognition via Action-Appearance Aligned Meta-Adaptation 无监督元学习算法
7. MGSampler: An Explainable Sampling Strategy for Video Action Recognition

### Video分割

1. Deep Transport Network for Unsupervised Video Object Segmentation
2. Graph Constrained Data Representation Learning for Human Motion Segmentation 人体动作分割
3. Joint Inductive and Transductive Learning for Video Object Segmentation 半监督视频物体分割
4. Domain Adaptive Video Segmentation via Temporal Consistency Regularization 以前的方法应用于不同领域性能明显下降，DA网络
5. Generating Masks From Boxes by Mining Spatio-Temporal Consistencies in Videos bbox生成分割，减少数据标注成本，增加泛化
6. Full-Duplex Strategy for Video Object Segmentation 外观和运动关系交叉注意模块，双向更新
7. Self-Supervised Video Object Segmentation by Motion Grouping 只使用光流信息，利用运动线索分割
8. End-to-End Video Instance Segmentation via Spatial-Temporal Graph Neural Networks 代表实例特征的图节点用于检测和分割，边用于跟踪，联合优化
9. Warp-Refine Propagation: Semi-Supervised Auto-Labeling via Cycle-Consistency 自动注释视频，改进的标签传播方法
10. Hierarchical Memory Matching Network for Video Object Segmentation 基于memory的方法(感觉需要了解一下memory相关的内容)
11. Video Instance Segmentation With a Propose-Reduce Paradigm 以前方法通常先对一帧或一个片段分割再跟踪或匹配导致错误累积，提出新范式
12. Weakly-Supervised Action Segmentation and Alignment via Transcript-Aware Union-of-Subspaces Learning 弱注释视频中分割动作
13. **Learning Motion-Appearance Co-Attention for Zero-Shot Video Object Segmentation** 外观和运动信息有效交互以适应复杂场景
14. Video Object Segmentation With Dynamic Memory Networks and Adaptive Object Alignment 改进之前方法的缺点，半监督
15. Unidentified Video Objects: A Benchmark for Dense, Open-World Segmentation 类似开放集的视频分割数据集
16. Learning To Better Segment Objects From Unseen Classes With Unlabeled Videos 如何使用未标记视频为不可见类对象自动生成训练数据
17. Refining Action Segmentation With Hierarchical Video Representations
18. Crossover Learning for Fast Online Video Instance Segmentation

### Video检测

1. Temporal Cue Guided Video Highlight Detection With Low-Rank Audio-Visual Fusion
2. PR-Net: Preference Reasoning for Personalized Video Highlight Detection 基于用户喜好个性化检测
3. TF-Blender: Temporal Feature Blender for Video Object Detection 视频目标检测，建模loe-level的时间关系增加特征表示
4. Temporal Action Detection With Multi-Level Supervision 半监督时间动作检测，改进动作不完整性带来的错误
5. Watch Only Once: An End-to-End Video Action Detection Framework 视频动作检测，涉及到定位和动作分类
6. HighlightMe: Detecting Highlights From Human-Centric Videos 领域和用户偏好无关的方法

## Video识别

1. Adaptive Focus for Efficient Video Recognition 空间冗余在视频识别中的应用，提高计算效率、易扩展
2. AdaMML: Adaptive Multi-Modal Learning for Efficient Video Recognition
3. TAM: Temporal Adaptive Module for Video Recognition 捕捉视频数据的不同运动模式
4. Condensing a Sequence to One Informative Frame for Video Recognition 将视频序列压缩成信息丰富的帧
5. Interactive Prototype Learning for Egocentric Action Recognition以自我为中心的视频识别，利用actor的运动线索进行学习
6. Efficient Action Recognition via Dynamic Knowledge Propagation 充分利用所选帧，交叉注意力机制动态知识传播，学生-教师框架
7. Searching for Two-Stream Models in Multivariate Space for Video Recognition 意识到之前工作手工设计两个stream和融合模块的缺点，利用NAS，改进架构
8. ORBIT: A Real-World Few-Shot Dataset for Teachable Object Recognition few-shot数据集
9. VideoLT: Large-Scale Long-Tailed Video Recognition 长尾视频识别数据集
10. Gait Recognition via Effective Global-Local Feature Representation and Local Temporal Aggregation 步态识别
11. Context-Sensitive Temporal Feature Learning for Gait Recognition
12. 3D Local Convolutional Neural Networks for Gait Recognition
13. Motion-Augmented Self-Training for Video Recognition at Smaller Scale

### Video插帧

1. Motion-Aware Dynamic Architecture for Efficient Frame Interpolation 为不同区域分配不同计算量，显著减少计算成本
2. Training Weakly Supervised Video Frame Interpolation With Events 基于事件的视频插针，引入注意力机制的弱监督
3. Asymmetric Bilateral Motion Estimation for Video Frame Interpolation
4. XVFI: eXtreme Video Frame Interpolation 提供了一个新的数据集和baseline

### Video对齐

找到一对视频之间的密集的时间对应

1. Normalized Human Pose Features for Human Action Video Alignment 只捕捉动作姿态特征，不受其他因素影响
2. TACo: Token-Aware Cascade Contrastive Learning for Video-Text Alignment 改进的对比学习用于视频文本对齐

### Video QA

1. Adversarial VQA: A New Benchmark for Evaluating the Robustness of VQA Models 提出了一个新的对抗性VQA benchmark
2. Env-QA: A Video Question Answering Benchmark for Comprehensive Understanding of Dynamic Environments
3. Unshuffling Data for Improved Generalization in Visual Question Answering 研究如何更好地泛化
4. TRAR: Routing the Attention Spans in Transformer for Visual Question Answering 动态调整transformer的全局和局部依赖建模
5. Greedy Gradient Ensemble for Robust Visual Question Answering 去偏框架，更好利用视觉信息
6. On the Hidden Treasure of Dialog in Video Question Answering 将对话转换为文本描述，比提供额外信息的工作表现更好
7. Just Ask: Learning To Answer Questions From Millions of Narrated Videos 不要人工注释，自动跨模态监督生成数据集
8. HAIR: Hierarchical Visual-Semantic Relational Reasoning for Video Question Answering 图记忆机制进行关系推理
9. Contrast and Classify: Training Robust VQA Models
10. Pano-AVQA: Grounded Audio-Visual Question Answering on 360deg Videos
11. Video Question Answering Using Language-Guided Deep Compressed-Domain Video Feature 开发了一种新的深度神经网络来提供从编码视频比特流获得的视频质量保证特征，以降低复杂性

### Video动作定位

1. Cross-Sentence Temporal and Semantic Relations in Video Activity Localisation
2. Foreground-Action Consistency Network for Weakly Supervised Temporal Action Localization 弱监督，保证前景与动作的一致性，又引入混合注意力机制
3. Learning Action Completeness From Points for Weakly-Supervised Temporal Action Localization 标签稀疏性导致问题，新框架生成密集伪标签并引入新loss
4. Video Self-Stitching Graph Network for Temporal Action Localization 针对短时间尺度的动作做出改进
5. Enriching Local and Global Contexts for Temporal Action Localization 主流pipeline的改进，丰富局部和全局上下文
6. Boundary-Sensitive Pre-Training for Temporal Localization in Videos
7. Divide and Conquer for Single-Frame Temporal Action Localization

### Video质量评估

1. Group-Aware Contrastive Regression for Action Quality Assessment
2. Unsupervised Curriculum Domain Adaptation for No-Reference Video Quality Assessment CNN泛化性能差，无监督DA处理跨域无参考VQA

### Video自然语言查询定位

自然语言查询定位视频中的时间段或帧

1. Support-Set Based Cross-Supervision for Video Grounding
2. Fast Video Moment Retrieval 旨在通过给定的自然语言句子对目标moment进行高效、准确的定位
3. Zero-Shot Natural Language Video Localization 消除注释成本
4. STVGBert: A Visual-Linguistic Transformer Based Framework for Spatio-Temporal Video Grounding 第一个单阶段方法，同时在空间和时间域定位目标
5. Universal Representation Learning From Multiple Domains for Few-Shot Classification

### Video去噪、去模糊

1. Patch Craft: Video Denoising by Deep Modeling and Patch Matching 提出的算法对CNN的输入进行处理，提高现有方法的表现
2. Unsupervised Deep Video Denoising
3. Bringing Events Into Video Deblurring With Non-Consecutively Blurry Frames 高时间速率的事件有助于去模糊
4. Multi-Scale Separable Network for Ultra-High-Definition Video Deblurring

### Video超分辨率

1. Omniscient Video Super-Resolution 新的框架，之前的框架可以视为该框架特殊情况
2. COMISR: Compression-Informed Video Super-Resolution 考虑压缩导进行高分辨率恢复
3. Deep Blind Video Super-Resolution
4. Real-World Video Super-Resolution: A Benchmark Dataset and a Decomposition Based Learning Scheme 视频超分辨率数据集和benckmark

### Video生成

1. iPOKE: Poking a Still Image for Controlled Stochastic Video Synthesis 针对视频中的动作作出的改进
2. Click To Move: Controlling Video Generation With Sparse Motion 用户可以通过鼠标点击指定场景中关键对象的简单对象轨迹来控制合成视频的运动

### Video异常检测

1. Exploring Temporal Coherence for More General Video Face Forgery Detection 伪造人脸视频检测
2. ID-Reveal: Identity-Aware DeepFake Video Detection DeepFake伪造检测，针对之前工作泛化性能较差做出改进
3. Dance With Self-Attention: A New Look of Conditional Random Fields on Anomaly Detection in Videos
4. A Hybrid Video Anomaly Detection Framework via Memory-Augmented Flow Reconstruction and Flow-Guided Frame Prediction

### Video inpainting

用与视频其余部分自然融合的内容填充时空区域

1. Flow-Guided Video Inpainting With Scene Templates 场景图像生成模型和场景到图像的映射
2. Frequency-Aware Spatiotemporal Transformers for Video Inpainting Detection 检测Video inpainting，从空间、时间和频率域共同挖掘视频修复的痕迹

### Video Entailment

确定假设文本陈述是否与前提视频相蕴涵或相矛盾

1. Explainable Video Entailment With Grounded Visual Evidence

### 视频文本检索

1. HiT: Hierarchical Transformer With Momentum Contrast for Video-Text Retrieval 动量跨模态交互对比

### 视频预测

1. A Hierarchical Variational Neural Uncertainty Model for Stochastic Video Prediction

### 主动speaker检测

1. Move2Hear: Active Audio-Visual Source Separation 主动音频视频源分离问题，同时多音频源分离目标对象的声音
2. MAAS: Multi-Modal Assignation for Active Speaker Detection 新方法，解决主动说话人检测问题的多模态性质
3. The Right To Talk: An Audio-Visual Transformer Approach

### 视频编码

1. ELF-VC: Efficient Learned Flexible-Rate Video Coding

### 视频表征学习

1. Motion-Focused Contrastive Learning of Video Representations 视频表征对比学习，关注运动
2. CrossCLR: Cross-Modal Contrastive Learning for Multi-Modal Video Representations 跨模态视频表征对比学习，提出了一个loss和嵌入
3. ASCNet: Self-Supervised Video Representation Learning With Appearance-Speed Consistency 提出正样本之间的一致性是学习鲁棒视频表示的关键
4. Self-Supervised Video Representation Learning With Meta-Contrastive Network
5. Multiview Pseudo-Labeling for Semi-Supervised Learning From Video

### 其他

1. Social Fabric: Tubelet Compositions for Video Relation Detection 视频理解，识别对象间主谓宾关系
2. Self-Conditioned Probabilistic Learning of Video Rescaling 学习保留更多信息的缩放视频
3. Hybrid Neural Fusion for Full-Frame Video Stabilization 混合空间融合，稳定视频，减轻光流不准确和伪影
4. Rethinking Self-Supervised Correspondence Learning: A Video Frame-Level Similarity Perspective 使用视频帧级的相关性来学习时空对应关系
5. Online-Trained Upsampler for Deep Low Complexity Video Compression 上采样器
6. BlockCopy: High-Resolution Video Processing With Block-Sparse Feature Propagation and Online Policies 根据图像只有很少的区域是重要的认识，使用稀疏卷积和时间特征传播来加快速度
7. Multiple Pairwise Ranking Networks for Personalized Video Summarization 视频摘要，任务表现受用户偏好影响，使用预定义的分类用户标签训练
8. Motion Guided Region Message Passing for Video Captioning 视频字幕生成，通过运动引导的跨帧消息传递感知时间结构
9. Long Short View Feature Decomposition via Contrastive Video Representation Learning 静态属性和非静态属性学习
10. TeachText: CrossModal Generalized Distillation for Text-Video Retrieval 文本视频检索
11. DRIVE: Deep Reinforced Accident Anticipation With Visual Explanation 交通事故预测引入带有注意力机制的视觉解释
12. ODAM: Object Detection, Association, and Mapping Using Posed RGB Video
13. Video Autoencoder: Self-Supervised Disentanglement of Static 3D Structure and Motion 学习三维结构和相机姿态的分离表示
14. Extending Neural P-Frame Codecs for B-Frame Coding
15. Out-of-Boundary View Synthesis Towards Full-Frame Video Stabilization
16. **Multimodal Clustering Networks for Self-Supervised Learning From Unlabeled Videos 多模态自监督视频学习**
17. Anonymizing Egocentric Videos 针对可能对以自我为中心的视频产生的攻击，研究如何添加扰动保护隐私
18. VIL-100: A New Dataset and a Baseline Model for Video Instance Lane Detection 新的视频实例车道检测数据集
19. Neural Video Portrait Relighting in Real-Time via Consistency Modeling 视频肖像重新照明
20. A Light Stage on Every Desk 重新照明
21. STRIVE: Scene Text Replacement in Videos 视频中的文本替换
22. Motion Guided Attention Fusion To Recognize Interactions From Videos 细粒度交互识别
23. Detecting Human-Object Relationships in Videos 静态图像人-物关系检测的延伸
24. Attentive and Contrastive Learning for Joint Depth and Motion Field Estimation 从单目视频中估计三维物体运动场
25. Modelling Neighbor Relation in Joint Space-Time Graph for Video Correspondence Learning
26. What You Can Learn by Staring at a Blank Wall 从观察空白墙的视频来推断人数或一个人的活动
27. Broaden Your Views for Self-Supervised Video Learning 自监督视频学习框架
28. Detecting Invisible People 单目视频检测和跟踪
29. On Compositions of Transformations in Contrastive Self-Supervised Learning
30. Procedure Planning in Instructional Videos via Contextual Modeling and Model-Based Policy Learning
31. Video Matting via Consistency-Regularized Graph Neural Networks
32. Overfitting the Data: Compact Neural Video Delivery via Content-Aware Feature Modulation 视频传输
33. Video Geo-Localization Employing Geo-Temporal Feature Learning and GPS Trajectory Smoothing 视频地理定位
34. ACAV100M: Automatic Curation of Large-Scale Datasets for Audio-Visual Video Representation Learning 视听对应数据集，非常适合自监督视频表征学习
35. Frozen in Time: A Joint Video and Image Encoder for End-to-End Retrieval 视频文本检索

## AI视觉安全

1. Exploiting Explanations for Model Inversion Attacks
2. Black-Box Detection of Backdoor Attacks With Limited Information and Data
3. A Backdoor Attack Against 3D Point Cloud Classifiers 3D点云分类器的后门攻击
4. Knowledge-Enriched Distributional Model Inversion Attacks
5. Membership Inference Attacks Are Easier on Difficult Problems MIA:检测数据样本是否用于训练，提出一种新的可预测性误差
6. Adversarial Attacks Are Reversible With Natural Supervision 修改被攻击的图像恢复自然的结构能逆转许多类型的攻击
7. Removing Adversarial Noise in Class Activation Feature Space 在类激活映射的特征空间中实现一种自监督的对抗性训练机制消除对抗性噪声
8. Aha! Adaptive History-Driven Attack for Decision-Based Black-Box Models
9. Multi-Expert Adversarial Attack Detection in Person Re-Identification Using Context Inconsistency
10. Rethinking the Backdoor Attacks' Triggers: A Frequency Perspective
11. Data-Free Universal Adversarial Perturbation and Black-Box Attack
12. Attack As the Best Defense: Nullifying Image-to-Image Translation GANs via Limit-Aware Adversarial Attack
13. Sample Efficient Detection and Classification of Adversarial Attacks via Self-Supervised Embeddings
14. Invisible Backdoor Attack With Sample-Specific Triggers
15. TkML-AP: Adversarial Attacks to Top-k Multi-Label Learning
16. CLEAR: Clean-Up Sample-Targeted Backdoor in Neural Networks
17. Augmented Lagrangian Adversarial Attacks
18. Meta-Attack: Class-Agnostic and Model-Agnostic Physical Adversarial Attack
19. Detection and Continual Learning of Novel Face Presentation Attacks
20. Consistency-Sensitivity Guided Ensemble Black-Box Adversarial Attacks in Low-Dimensional Spaces
21. PointBA: Towards Backdoor Attacks in 3D Point Cloud
22. Practical Relative Order Attack in Deep Ranking
23. Defending Against Universal Adversarial Patches by Clipping Feature Norms
24. Just One Moment: Structural Vulnerability of Deep Action Recognition Against One Frame Attack
25. On Generating Transferable Targeted Perturbations
26. Admix: Enhancing the Transferability of Adversarial Attacks
27. LIRA: Learnable, Imperceptible and Robust Backdoor Attacks
28. Integer-Arithmetic-Only Certified Robustness for Quantized Neural Networks
29. Parallel Rectangle Flip Attack: A Query-Based Black-Box Attack Against Object Detection 针对目标检测的黑盒方法

## 场景图生成

1. Spatial-Temporal Transformer for Dynamic Scene Graph Generation
2. Grounding Consistency: Distilling Spatial Common Sense for Precise Visual Relationship Detection 探讨了之前方法做法的缺陷，从空间和视觉证据的角度进行改进
3. Self-Supervised Real-to-Sim Scene Generation 选了场景图生成作为下游任务，自监督合成数据
4. Segmentation-Grounded Scene Graph Generation 基于像素级分割的场景图生成框架
5. Topic Scene Graph Generation by Attention Distillation From Caption 用图像标题的概括信息指导场景图生成
6. Context-Aware Scene Graph Generation With Seq2Seq Transformers 显式限制已经预测的关系，自回归方式一次预测一个视觉关系，基于强化学习的训练策略
7. Target Adaptive Context Aggregation for Video Scene Graph Generation
8. Visual Distant Supervision for Scene Graph Generation 新的视觉关系学习范式，没有任何标记数据训练
9. A Simple Baseline for Weakly-Supervised Scene Graph Generation
10. Learning To Generate Scene Graph From Natural Language Supervision 图像文本对中学习提取图像中局部对象及其关系的图形表示方法
11. From General to Specific: Informative Scene Graph Generation via Balance Adjustment

## 医学相关

1. Long-Term Temporally Consistent Unpaired Video Translation From Simulated Surgical 3D Data 个人感觉本质还是在扩充医学数据上，可能是没理解什么是Video Translation
2. Mutual-Complementing Framework for Nuclei Detection and Segmentation in Pathology Image 检测和分割互补框架
3. The Way to My Heart Is Through Contrastive Learning: Remote Photoplethysmography From Unlabelled Video 从视频中估计生理信号，对比学习
4. CDNet: Centripetal Direction Network for Nuclear Instance Segmentation 细胞核实例分割
5. The Benefit of Distraction: Denoising Camera-Based Physiological Measurements Using Inverse Attention 使用反向注意力机制去噪
6. Collaborative and Adversarial Learning of Focused and Dispersive Representations for Semi-Supervised Polyp Segmentation 结肠镜图像中分割息肉，半监督，辅助对抗学习
7. Visual-Textual Attentive Semantic Consistency for Medical Report Generation 自动医学报告生成，多模态，X射线输出文本报告
8. GLoRIA: A Multimodal Global-Local Representation Learning Framework for Label-Efficient Medical Image Recognition
9. Sparse-Shot Learning With Exclusive Cross-Entropy for Extremely Many Localisations
10. Self-Supervised Vessel Segmentation via Adversarial Learning
11. Cortical Surface Shape Analysis Based on Alexandrov Polyhedra
12. CCT-Net: Category-Invariant Cross-Domain Transfer for Medical Single-to-Multiple Disease Diagnosis
13. T-AutoML: Automated Machine Learning for Lesion Segmentation Using Transformers in 3D Medical Imaging
14. Big Self-Supervised Models Advance Medical Image Classification 自监督预训练在医学图像分类的有效性
15. Recurrent Mask Refinement for Few-Shot Medical Image Segmentation

## 深度估计、预测

1. Bayesian Deep Basis Fitting for Depth Completion With Uncertainty
2. StructDepth: Leveraging the Structural Regularities for Self-Supervised Indoor Depth Estimation 自监督 室内 单目深度估计
3. Bridging Unsupervised and Supervised Depth From Focus via All-in-Focus Supervision 可以无监督和监督地训练
4. R-MSFM: Recurrent Multi-Scale Feature Modulation for Monocular Depth Estimating 单目深度估计，递归多尺度特征，改进了Unet方法的缺点
5. Towards Interpretable Deep Networks for Monocular Depth Estimation 可解释性更强的单目深度估计网络
6. Adaptive Confidence Thresholding for Monocular Depth Estimation 自监督立体匹配生成立体图像的伪地面深度图
7. Can Scale-Consistent Monocular Depth Be Learned in a Self-Supervised Scale-Invariant Manner? 提出一种新的loss，通过点云对齐来实现尺度一致，自监督
8. Revisiting Stereo Depth Estimation From a Sequence-to-Sequence Perspective With Transformers 立体深度估计，seq-to-seq重新建模
9. Augmenting Depth Estimation With Geospatial Context 使用地理空间上下文的深度估计
10. Self-Supervised Monocular Depth Estimation for All Day Images Using Domain Separation 白天黑夜存在较大的域便宜和光照变化，提出全天图像自监督深度估计域分离网络
11. Adaptive Surface Normal Constraint for Depth Estimation
12. Transformer-Based Attention Networks for Continuous Pixel-Wise Prediction
13. Fine-Grained Semantics-Aware Representation Enhancement for Self-Supervised Monocular Depth Estimation
14. Unsupervised Depth Completion With Calibrated Backprojection Layers 图像和稀疏点云推断密集深度
15. DnD: Dense Depth Estimation in Crowded Dynamic Indoor Scenes
16. Revealing the Reciprocal Relations Between Self-Supervised Stereo and Monocular Depth Estimation 自监督，利用单目和立体两个任务的互补知识来改进性能
17. Boosting Monocular Depth Estimation With Lightweight 3D Point Fusion 通过添加3D点作为深度引导来增强单目深度估计
18. DepthInSpace: Exploitation and Fusion of Multiple Video Frames for Structured-Light Depth Estimation

## 手语相关

1. Mixed SIGNals: Sign Language Production via a Mixture of Motion Primitives 口语生成连续的手语动画序列
2. SignBERT: Pre-Training of Hand-Model-Aware Representation for Sign Language Recognition 自监督带有手部先验
3. Self-Mutual Distillation Learning for Continuous Sign Language Recognition 知识蒸馏增强模块的识别能力
4. Visual Alignment Constraint for Continuous Sign Language Recognition 强化特征提取部分，缓解过拟合
5. Stochastic Transformer Networks With Linear Competing Units: Application To End-to-End SL Translation 自动手语翻译，只需要文本GT？
6. Aligning Subtitles in Sign Language Videos

## 强化学习

1. Partial Off-Policy Learning: Balance Accuracy and Diversity for Human-Oriented Image Captioning 面向人的图像字幕，设计一个新的策略
2. MEDIRL: Predicting the Visual Attention of Drivers via Maximum Entropy Deep Inverse Reinforcement Learning 反向强化学习预测事故多发情况下驾驶员的视觉注意，新的驾驶员注意力数据集
3. GridToPix: Training Embodied Agents With Minimal Supervision 深度强化学习奖励机制研究

## 联邦学习

1. Federated Learning for Non-IID Data via Unified Feature Learning and Optimization Objective Alignment 非独立同分布下的联邦学习
2. Collaborative Unsupervised Visual Representation Learning From Decentralized Data
3. Ensemble Attention Distillation for Privacy-Preserving Federated Learning

## 自动驾驶

1. NEAT: Neural Attention Fields for End-to-End Autonomous Driving
2. LookOut: Diverse Multi-Future Prediction and Planning for Self-Driving 多样性运动预测
3. Prediction by Anticipation: An Action-Conditional Prediction Method Based on Interaction Learning 将交互作用视为潜在的概率生成过程，通过预期进行预测
4. TMCOSS: Thresholded Multi-Criteria Online Subset Selection for Data-Efficient Autonomous Driving 高效利用数据进行训练
5. Large Scale Interactive Motion Forecasting for Autonomous Driving: The Waymo Open Motion Dataset
6. MultiSiam: Self-Supervised Multi-Instance Siamese Representation Learning for Autonomous Driving 利用未标记数据进行表示学习的自监督学习，改进现有方法的缺德
7. Road Anomaly Detection by Partial Image Reconstruction With Segmentation Coupling 自动驾驶环境下检测未知物体的新方法，描述为异常检测

## Object导航

1. Auxiliary Tasks and Exploration Enable ObjectGoal Navigation 在不可见的环境中进行到某个对象的导航，添加辅助任务和探索奖励进行改进
2. Hierarchical Object-to-Zone Graph for Object Navigation 分层obj-to-zone图（hoz），从粗到细引导agent，并且有在线学习机制，会更新hoz
3. The Road To Know-Where: An Object-and-Room Informed Sequential BERT for Indoor Vision-Language Navigation VLN导航
4. Pathdreamer: A World Model for Indoor Navigation VLN导航作为下游任务测试
5. Episodic Transformer for Vision-and-Language Navigation
6. RobustNav: Towards Benchmarking Robustness in Embodied Navigation
7. Self-Motivated Communication Agent for Real-World Vision-Dialog Navigation
8. Vision-Language Navigation With Random Environmental Mixup 数据增广减少小数据规模和大导航空间的差异

## 事件摄像机

1. Graph-Based Asynchronous Event Processing for Rapid Object Recognition 事件照相机的物体识别
2. EvIntSR-Net: Event Guided Multiple Latent Frames Reconstruction and Super-Resolution 事件摄像机重建高分辨率强度图像
3. The Spatio-Temporal Poisson Point Process: A Simple Model for the Alignment of Event Camera Data 事件数据自然时空结构捕捉
4. N-ImageNet: Towards Robust, Fine-Grained Object Recognition With Event Cameras 一个新的事件摄像机细粒度数据集
5. Event-Intensity Stereo: Estimating Depth by the Best of Both Worlds

## 人机交互检测

1. Spatially Conditioned Graphs for Detecting Human-Object Interactions
2. Exploiting Scene Graphs for Human-Object Interaction Detection 场景图进行人-物交互检测
3. Virtual Multi-Modality Self-Supervised Foreground Matting for Human-Object Interaction 人像抠图算法，学习人-物交互前景
4. Weakly Supervised Human-Object Interaction Detection in Video via Contrastive Spatiotemporal Regions 人机交互检测
5. Discovering Human Interactions With Large-Vocabulary Objects via Query and Multi-Scale Detection 大量对象类别下的人机交互检测

## 其他

1. Out-of-Core Surface Reconstruction via Global TGV Minimization 表面重建
2. Extensions of Karger's Algorithm: Why They Fail in Theory and How They Are Useful in Practice 扩展karger算法，半监督学习
3. Paint Transformer: Feed Forward Neural Painting With Stroke Prediction 神经绘画，以前是RL方法，这篇用Transformer
4. YouRefIt: Embodied Reference Understanding With Language and Gesture 不知道是不是新的任务，机器理解embodied reference
5. SACoD: Sensor Algorithm Co-Design Towards Efficient CNN-Powered Intelligent PhlatCam 硬件相关设计CNN系统
6. Audio2Gestures: Generating Diverse Gestures From Speech Audio With Conditional Variational Autoencoders 语音音频生成对话手势
7. Keep CALM and Improve Visual Feature Attribution 改进CAM
8. DocFormer: End-to-End Transformer for Document Understanding VDU 视觉电子文档理解
9. Towards Better Explanations of Class Activation Mapping 针对CAM的改进，也能算可解释性研究
10. Learn-To-Race: A Multimodal Control Environment for Autonomous Racing 提出了新的任务，学习自动赛车
11. Large Scale Multi-Illuminant (LSMI) Dataset for Developing White Balance Algorithm Under Mixed Illumination 大规模多光源数据集论文
12. Looking Here or There? Gaze Following in 360-Degree Images 360度图像视觉追踪，数据集论文，并提出了一种新的框架
13. DeepMultiCap: Performance Capture of Multiple Characters Using Sparse Multiview Cameras 稀疏多视角相机 human performance capture
14. AI Choreographer: Music Conditioned 3D Dance Generation With AIST++ 给了一个数据集，音乐生成三维运动，多模态
15. SOMA: Solving Optical Marker-Based MoCap Automatically 对基于光学标记的动作捕捉得到的带有噪音和不完整原始输出进行标记
16. Detector-Free Weakly Supervised Grounding by Separation 图像加自由形式的对应文本，学习突出文本对应的图像区域，可以理解为一种检测任务
17. Deep Virtual Markers for Articulated 3D Shapes 映射铰接3D的3D点到虚拟标记标签中，估计三维数据密集和准确的位置信息
18. M3D-VTON: A Monocular-to-3D Virtual Try-On Network 单目虚拟3D试穿
19. ZFlow: Gated Appearance Flow-Based Virtual Try-On With 3D Priors
20. FashionMirror: Co-Attention Feature-Remapping Virtual Try-On With Sequential Template Poses
21. Inferring High-Resolution Traffic Accident Risk Maps Based on Satellite Imagery and GPS Trajectories 高分辨率事故风险图
22. The Surprising Effectiveness of Visual Odometry Techniques for Embodied PointGoal Navigation 机器人导航
23. STR-GQN: Scene Representation and Rendering for Unknown Cameras Based on Spatial Transformation Routing 提出一种新的机制，据说效果比空间相关任务的几何感知模块好
24. Visual Scene Graphs for Audio Source Separation 视觉引导音频源分离
25. How To Design a Three-Stage Architecture for Audio-Visual Active Speaker Detection in the Wild 检测Active Speaker
26. Social NCE: Contrastive Learning of Socially-Aware Motion Representations 多智能体社会感知运动表示对比学习
27. FloW: A Dataset and Benchmark for Floating Waste Detection in Inland Waters 内陆水域漂浮垃圾检测的数据集和基准
28. VSAC: Efficient and Accurate Estimator for H and F
29. Fusion Moves for Graph Matching 二次分配问题近似算法
30. Fast Light-Field Disparity Estimation With Multi-Disparity-Scale Cost Aggregation 光场视差恢复
31. Faster Multi-Object Segmentation Using Parallel Quadratic Pseudo-Boolean Optimization 改进QPBO算法，加快运行速度
32. Interpretation of Emergent Communication in Heterogeneous Collaborative Embodied Agents
33. A Dark Flash Normal Camera 优化立体几何、填充图像阴影
34. Dynamic CT Reconstruction From Limited Views With Implicit Neural Representations and Parametric Motion Fields 4D-CT重建
35. 4D Cloud Scattering Tomography
36. De-Rendering Stylized Texts 注意到排版信息的文字转化
37. Channel Augmented Joint Learning for Visible-Infrared Recognition 红外识别问题
38. A Hybrid Frequency-Spatial Domain Model for Sparse Image Reconstruction in Scanning Transmission Electron Microscopy
39. Efficient Large Scale Inlier Voting for Geometric Vision Problems 离群点剔除算法
40. Act the Part: Learning Interaction Strategies for Articulated Object Part Discovery 感知与互动间关系的研究
41. Self-Supervised Cryo-Electron Tomography Volumetric Image Restoration From Single Noisy Volume With Sparsity Constraint 从单个噪声数据中进行恢复
42. Weakly Supervised Temporal Anomaly Segmentation With Dynamic Time Warping 从时态数据中检测异常，实例级粗粒度标签训练输出分割结果
43. Unsupervised Dense Deformation Embedding Network for Template-Free Shape Correspondence 可变形3D形状对齐
44. Learning Indoor Inverse Rendering With 3D Spatially-Varying Lighting 从单个图像联合估计法线、深度和3D空间变化照明
45. Learning To Drive From a World on Rails 从预先记录的架势日志中学习基于视觉的交互式驾驶策略，有world on rails的假设
46. FFT-OT: A Fast Algorithm for Optimal Transportation 最优交通图求解
47. TRiPOD: Human Trajectory and Pose Dynamics Forecasting in the Wild 人体轨迹和姿态动力学联合预测
48. LapsCore: Language-Guided Person Search via Color Reasoning 语言引导的人物搜索，从庞大的图像数据库中检索与给的文本查询最匹配的人
49. H2O: A Benchmark for Visual Human-Human Object Handover Analysis 人机对象切换的可视化分析数据集
50. Beyond Road Extraction: A Dataset for Map Update Using Aerial Images 从道路提取过渡到地图更新任务，提出了一个新的数据集
51. Transfusion: A Novel SLAM Method Focused on Transparent Objects 改进RGB-D SLAM算法，允许透明物体存在和恢复
52. iMAP: Implicit Mapping and Positioning in Real-Time
53. Multi-Modality Associative Bridging Through Memory: Speech Sound Recollected From Face Video 视听多模态桥接框架，文中用于两个下游任务：唇读和无声视频语音重建
54. Where2Act: From Pixels to Actions for Articulated 3D Objects 研究感知可交互的物体
55. CAPTRA: CAtegory-Level Pose Tracking for Rigid and Articulated Objects From Point Clouds
56. Toward Human-Like Grasp: Dexterous Grasping via Semantic Representation of Object-Hand 机器人手抓取数据集
57. Hand-Object Contact Consistency Reasoning for Human Grasps Generation
58. Graspness Discovery in Clutters for Fast and Accurate Grasp Detection 抓取姿态检测
59. Single Image 3D Shape Retrieval via Cross-Modal Instance and Category Contrastive Learning 单个图像的3D形状检索
60. Omnidata: A Scalable Pipeline for Making Multi-Task Mid-Level Vision Datasets From 3D Scans 提供了一个生成数据集的pipeline（我们研究了几个使用这个过程来分析数据、任务、模型和学习算法之间关系的小例子，并提出了一些超出本文范围的令人兴奋的方向。）
61. Hypersim: A Photorealistic Synthetic Dataset for Holistic Indoor Scene Understanding 数据集论文
62. DWKS: A Local Descriptor of Deformations Between Meshes and Point Clouds 新的pointwise描述，寻找两个可变形形状集合之间的对应关系
63. FloorPlanCAD: A Large-Scale CAD Drawing Dataset for Panoptic Symbol Spotting 数据集论文
64. Event-Based Video Reconstruction Using Transformer 基于事件照相机的视频重建
65. DiagViB-6: A Diagnostic Benchmark Suite for Vision Models in the Presence of Shortcut and Generalization Opportunities 数据集和指标
66. Super Resolve Dynamic Scene From Continuous Spike Streams
67. Product1M: Towards Weakly Supervised Instance-Level Product Retrieval via Cross-Modal Pretraining 实例级检索数据集
68. Factorizing Perception and Policy for Interactive Instruction Following
69. 4DComplete: Non-Rigid Motion Estimation Beyond the Observable Surface 非刚性4D重建和数据集
70. E-ViL: A Dataset and Benchmark for Natural Language Explanations in Vision-Language Tasks 自然语言解释数据集
71. A New Journey From SDRTV to HDRTV SDR内容到HDR的数据集、指标和基线
72. ELLIPSDF: Joint Object Pose and Shape Optimization With a Bi-Level Ellipsoid and Signed Distance Function Description 基于RGB-D的SLAM
73. Product Quantizer Aware Inverted Index for Scalable Nearest Neighbor Search
74. On Exposing the Challenging Long Tail in Future Prediction of Traffic Actors 交通场景下的未来预测
75. ICON: Learning Regular Maps Through Inverse Consistency 学习数据样本之间的映射
76. FLAR: A Unified Prototype Framework for Few-Sample Lifelong Active Recognition 主动识别
77. VLGrammar: Grounded Grammar Induction of Vision and Language 提出了一个新的数据集，共同归纳视觉语法和语言语法的框架
78. BuildingNet: Learning To Label 3D Buildings 3D建筑模型数据集
79. Q-Match: Iterative Shape Matching via Quantum Annealing 寻找形状对应关系
80. Cross-Encoder for Unsupervised Gaze Representation Learning 3D凝视估计检测
81. H2O: Two Hands Manipulating Objects for First Person Interaction Recognition 新的数据集，双手3D姿势和被操纵对象6D姿势注释，预测交互类别
82. Infinite Nature: Perpetual View Generation of Natural Scenes From a Single Image 永久视图生成，提出一个新的数据集
83. CVIS的2D/3D车辆解析，提出一种多视图数据合成算法，一种2D/3D车辆解析算法
84. R-SLAM: Optimizing Eye Tracking From Rolling Shutter Video of the Retina 眼动追踪
85. End-to-End Trainable Trident Person Search Network Using Adaptive Gradient Propagation 人员搜索任务
86. GeomNet: A Neural Network Based on Riemannian Geometries of SPD Matrix Space and Cholesky Space for 3D Skeleton-Based Interaction Recognition 3D骨架互动识别