

结合GAN的零次学习 (zero-shot learning)

原创 Nango明楠 机器学习与生成对抗网络 今天

点击上方“机器学习与生成对抗网络”，关注“星标”

获取有趣、好玩的前沿干货！

零次学习 (zero-shot learning)

- 众所周知，深度学习的崛起依赖于大量的训练样本；监督式学习已经在各项任务上取得了极好的效果。
- 但有一点和我们人的“智能”不一样的是，一个两岁稚子第一次进动物园，看到老虎时，由于在家中见过猫，根据其父的描述“和猫很像、但比猫更大、有黑白黄斑纹、长牙等是老虎”即可轻易判断所见动物为老虎。
- 深度学习如果从来没有“见”过老虎图像样本，是很难去分类老虎的；而零次学习ZSL正是以解决此问题而提出：即便没见过老虎，但知道一些属性，例如黑白黄斑纹、尾巴、长牙等，通过这些语义属性即可将“未见类”识别出来。
- 例如对于图像分类中的ZSL，一般定义是，给定训练集 X_{train} （例如，见过猫的样子）、训练标签 Y_{train} （例如，标签是猫）、训练类别的（语义属性）描述 A_{train} （例如，比人小、四肢尾巴、有可爱的头）、测试标签 Y_{test} （例如，老虎）和测试类别的描述 A_{test} （例如，和猫类似，比较很大只，凶猛），训练一个分类器 C 。
- 测试时对于未见过（unseen）的 X_{test} （例如，动物园的老虎），分类器 C 可进行 Y_{test} 分类（例如，它是一只老虎！）。

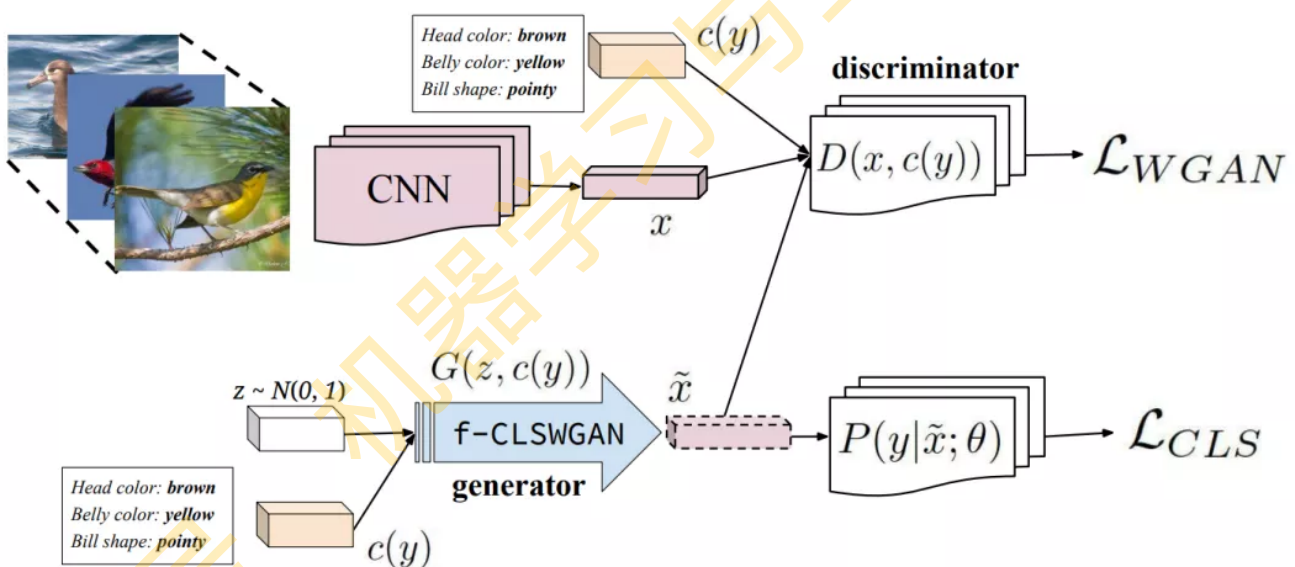
目录

- 2018_CVPR_FGN用GAN生成视觉特征_增加辅助分类引导_增强视觉特征的方式_可嵌合多种方法
- 2019_CVPR_Oral_GMN梯度匹配GAN_生成样本进行训练
- 2019_CVPR_ZS-SBIR_草图检索_类似GAN思想用GRL反转梯度让投影空间一致_TripletLoss_Attention特征图加权_提出新草图数据集
- 2019_CVPR_SEM-PCYC_草图检索_GAN让投影空间一致_分别重构损失_Autoencoder组合额外监督信息
- 2019_CVPR_LisGAN类别灵魂元向量和对应属性生成视觉向量的表示_GAN训练
- 2019_CVPR_CEWGAN-OD_特征熵分出seen和unseen_GAN生成unseen_重构损失_生成特征类别匹配损失

- 2019_ICCV_通过GAN生成unseen_参数化熵的测度_引导创造力生成
- 2019_ICCV_条件耦合GAN捕获不同任务的联合分布_域自适应Zeroshot
- 2019_ICCV_情感结构嵌入_首次识别未知情感Zeroshot_中间语义层GAN

2018_CVPR_FGN用GAN生成视觉特征_增加辅助分类引导_增强视觉特征的方式_可嵌合多种方法

- Feature Generating Networks for Zero-Shot Learning
- (https://zpascal.net/cvpr2018/Xian_Feature_Generating_Networks_CVPR_2018_paper.pdf)
- <https://www.mpi-inf.mpg.de/departments/computer-vision-and-multimodal-computing/research/zero-shot-learning/feature-generating-networks-for-zero-shot-learning/>



- 增强视觉特征的方法.可联合多种经典方法进行提高ZSL问题.
- 相比于生成图像, GAN直接在视觉空间上生成新特征(数据)会有效得多.
- 添加了auxiliary classifier, 让学习的特征具有强的discrimination.
- 带LeakyReLU的MLP来作为G和D网络, output是ReLU激活.
- G的输入: 类的语义向量(attribute),和服从高斯分布的随机值. 输出该类的虚拟视觉特征.
- 分类网络的输入: G输出的视觉特征, 指导让G生成的特征更具有区分性.
- 识别方式两种:
 - 通过G生成unseen和seen类的(x' , y)的数据集(视觉特征和类别), 从而训练这么个分类器来分类所有类别.

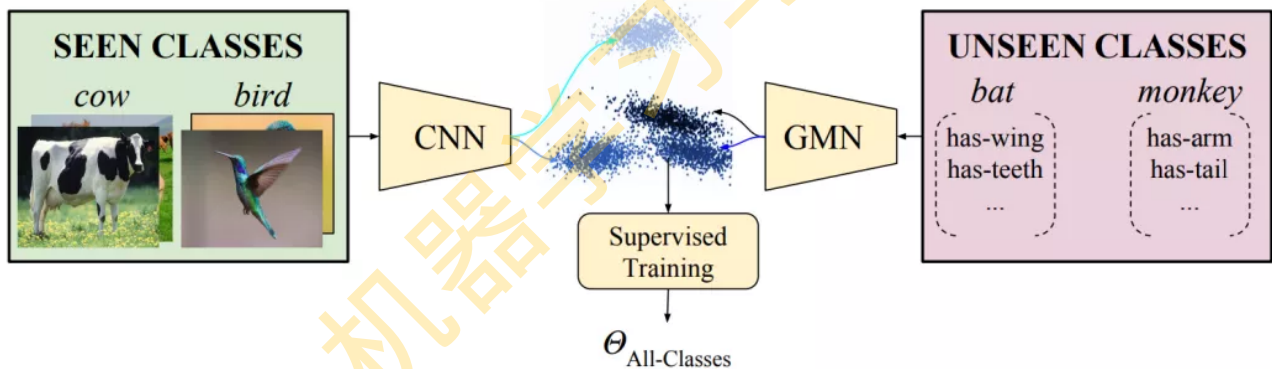
- 通过unseen的样本得到视觉特征 x , 再来判断属于哪一类(普通的分类方法)
- 通过 $x, attr(y)$ 来判断属于哪一类.
- 很多经典方法都可作为此处的model
- multimodal embedding model
- softmax classification

2019_CVPR_Oral_GMN梯度匹配GAN_生成样本进行训练

- 2019_Oral_Gradient Matching Generative Networks for Zero-Shot Learning
- http://openaccess.thecvf.com/content_CVPR_2019/papers/Sariyildiz_Gradient_Matching_Generative_Networks_for_Zero-Shot_Learning_CVPR_2019_paper.pdf

采用GAN生成样本(生成seen和unseen样本), 提出新的Loss来测量生成样本质量(通过梯度信号), 促使生成样本更为真实.

本文学习一个生成模型, 生成 seen和unseen 的样本, 并让生成样本和seen样本作为训练集来训练分类器, 从而将zeroshot问题转换成监督学习问题.



有几种方式可让类别之间有关联关系:

- (1) 手动定义物体的视觉和语义特性的属性;
- (2) 根据类名获得词向量;
- (3) 从类别关联树(relevant taxonomy tree)上构建向量.

ZSL问题可以归结为: 对视觉特征空间(如pretrained CNN提取的特征) 与类别嵌入空间 进行的关系建模

与其他 GAN 样本生成样本一样, 最终分类准确性很大程度取决于GAN所生成样本的多样性和真实性.

本文关注:

- (1) 利用生成样本隐式建模每个unseen类的流形;
- (2) 确保用生成本来训练的分类器效果更好.

本文采用无监督GAN, 提出了新的损失函数(检查生成样本的质量), 促进满足上面两点要求.

且文中采用unconditional GAN 鉴别器, 不依赖于语义嵌入向量, 有助于 unseen样本的生成.

对于条件生成模型, 监督信号是至关重要的, 以往采用分类Loss来衡量seen的准确性. 但本文认为分类Loss不一定会得到好的训练集(seen+ 合成样本), 因为当 G 生成易分类的样本时, 分类Loss会特别低, 从而导致生成的样本质量不够高. 本文希望生成 '能让模型训练得更好' 的样本 (即高质量样本).

文中观察到: 一个生成模型在学习真实类别的流行时, 在生成样本集合上训练的分类模型参数的偏导数 和在真实样本集合上训练的分类模型参数的偏导数 是高度相关的. 故提出**梯度匹配 Loss (gradient matching loss)** 用于测量 真实样本和合成样本所获得的梯度向量的差异, 让两者偏导数相近.

之后, 对于生成的样本和真实seen样本构成训练集, 此时所有类别均出现在训练集中, 故采用普通的多分类方法即可.

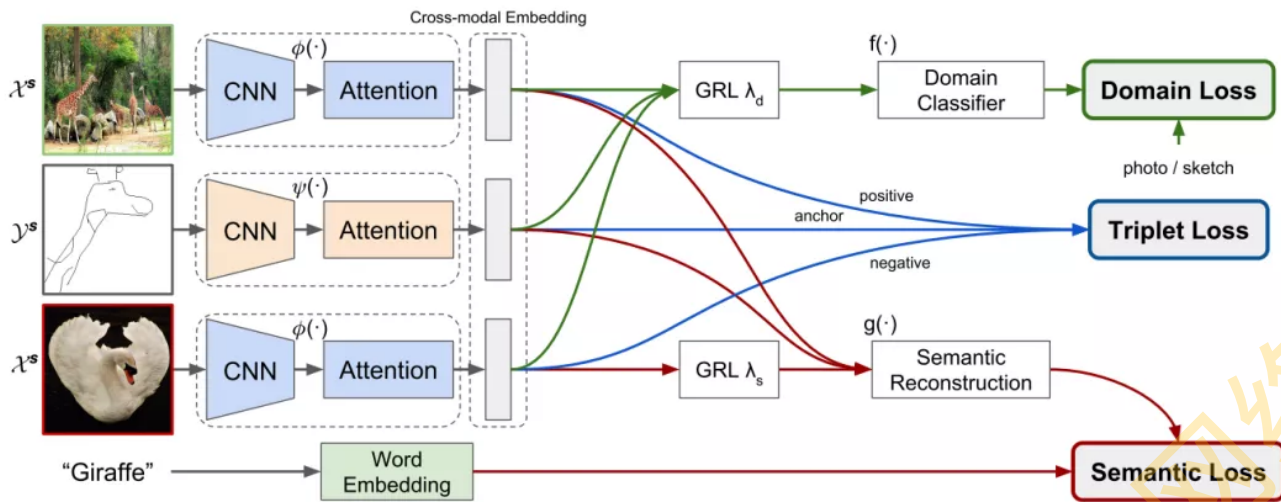
2019_CVPR_ZS-SBIR_草图检索_类似GAN思想用GRL反转梯度让投影空间一致_TripletLoss_Attention特征图加权_提出新草图数据集

- 2019_Doodle to Search Practical Zero-Shot Sketch-based Image Retrieval
- http://openaccess.thecvf.com/content_CVPR_2019/papers/Dey_Doodle_to_Search_Practical_Zero-Shot_Sketch-Based_Image_Retrieval_CVPR_2019_paper.pdf

贡献了数据集: **QuickDraw-Extended**, 包含 330k 张 sketch草图以及 204k 张正常图片(共 110类). 适合用于做 **zero-shot 的图像检索**, zero-shot sketch-based image retrieval (ZS-SBIR). 该草图数据集比其他的数据集更加抽象(让domain gap更显著), 草图来自不同人绘制, 保证风格多样性.

提出该数据集是因为目前: 1) sketch 与image 的 domain gap 不够大. 即现有的草图有些太逼真; 2) sketch 需要高度抽象; 3) ZSL 可以从中进行语义迁移.

文中提出更好的 cross-domain transfer model: (1)提出领域缠绕模式让模型学出 领域无关 的嵌入向量. 且(2)且提出了好的 semantic loss 来确保语义信息的保留.



• Attention 结构:

- 是用 图像字幕的 soft-attention 方式, 来帮助网络定位重要的特征
- 学习出 feature mask 为 att 来对特征图进行加权,
- $f = f + f \times att$, 文中采用 1×1 的卷积+ mean方式获得 attention mask)

• 学习函数中有三个: Domain loss, Triplet loss, 以及semantic loss

• Triplet Loss (Ranking loss的一种):

- 草图的嵌入向量应该离同类图像的嵌入向量更近, 而和不同类的图像嵌入向量距离更远
- 标准的Triplet-loss做法, 计算向量间的 MSE 相似度, 同类的更近, 异类的更远.
- $loss_i = \max\{0, \mu + \|\psi(a) - \phi(p)\|_2 - \|\psi(a) - \phi(n)\|_2\}$

• Domain loss:

- Triplet-loss 不能保证将 sketch 和 image 映射到同一个空间中
- 采用论文2015_Unsupervised Domain Adaptation by Backpropagation(<http://202.38.196.91/cache/8/03/proceedings.mlr.press/07268e345891f67876df80b3ab011f88/ganin15.pdf>)方法, 让不同领域的样本, 他们投影后的向量空间一致.
- 类似对抗网络, domain分类器分辨"投影后的向量"是属于哪个领域(sketch还是image领域, 二分类). 而投影网络尽量让 domain分类损失最大(这样不同领域被投影后的向量, 都属于一个空间了)
- 但不是利用对抗网络来学习, 而是利用 反转梯度 来更新(从而使得domain loss的损失越来越大)
- 这种方法(GRL) 听说比GAN更好训练收敛.

• Semantic Loss:

- 通过语义信息, 重构语义向量.
- 通过投影后的嵌入向量, 重构出类别的语义向量(如词向量), 最小化重构和真实语义向量的cosine距离.
- 引入负样本, 一起来重构 正样本的语义向量.(即重构草图和正样本image的语义向量)

- 通过投影后的嵌入向量, 就可以进行检索任务,或零次学习任务了.
- 在baseline上加入attention的提升很大.

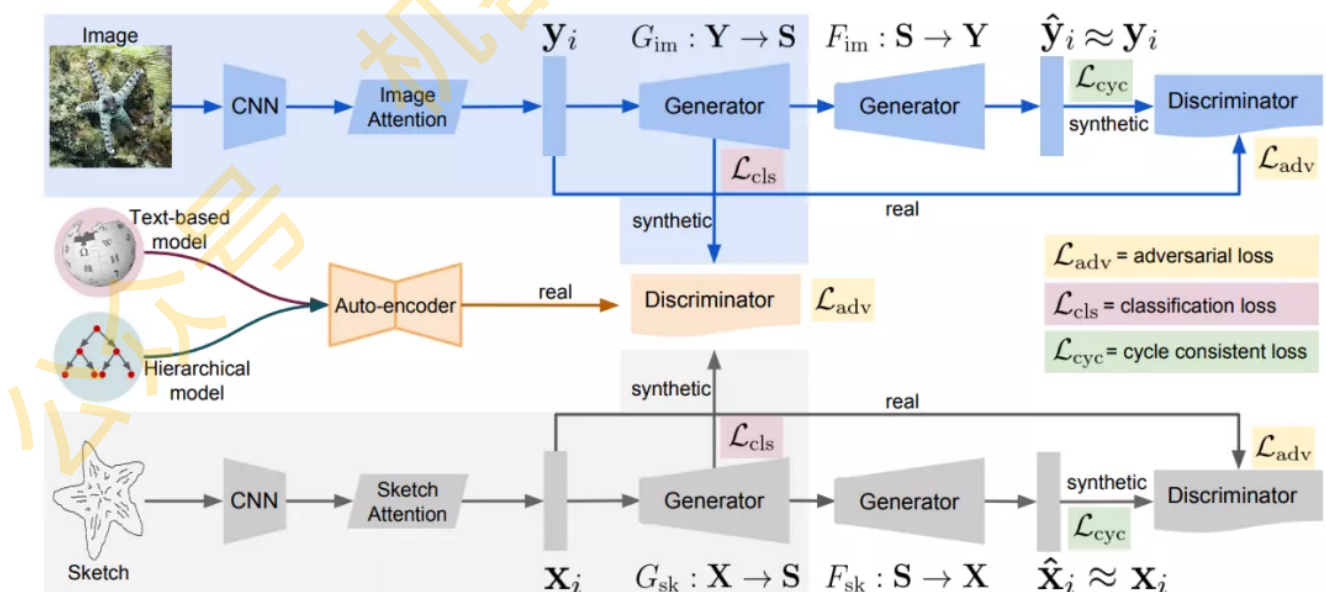
2019_CVPR_SEM-PCYC_草图检索_GAN让投影空间一致_分别重构损失_Autoencoder组合额外监督信息

- 2019_Semantically Tied Paired Cycle Consistency for Zero-Shot Sketch-based Image Retrieval
- <https://arxiv.org/pdf/1903.03372.pdf>

面临问题:

- (1) domain gap,
- (2) 同类的方差大,
- (3) 关于unseen的知识有限

- 提出semantically aligned paired cycle-consistent generative model(SEM-PCYC)方法
- 通过生成对抗学习将sketch-image映射到共同的语义空间, 通过cycle consistency从而不需要完全配对的 sketch-image 关系(即草图和image是对应同一个东西,比如同姿态同动作这些高要求).
- 通过特征选择导向的AutoEncoder来组合不同来源的额外信息(词向量, 结构树等), 从而选择能最小化类内方差,最大化类间方差的额外信息.



看图一目了然. 上面是image, 下面是草图sketch, 中间是额外知识引入.

- sketch和image映射到共同空间(domain gap):

- 从额外信息得鉴别器, 来判断image的生成器和sketch的生成器的输出的判断, 从而让sketch和image的投影向量属于同一空间(鉴别器难以分辨时说明处于同一空间)
- Cycle Consistency Loss:
 - 让sketch/image 的输出向量, 不仅能投影到同一语义向量中, 也能够映射回原先的特征空间中.
 - 即图中上下分支的右半部分. 通过重构模型实现.
- 分类损失:
 - 确保生成的语义向量具备类别可分类性, 从而能实现分类任务
 - 将sketch/image的输出向量(生成的投影向量)传入分类器.
- 模型的每个分支将sketch/image通过对抗学习 (共享语义判别器) 映射到共同语义空间,
- cycle consistency不需要对齐的sketch/image对, 即不需要 (sketch image)的内容是完全对应的配对关系. 且可在对抗训练中发挥正则作用.
- 分类损失确保sketch/image生成的伪语义向量的类别可区分性
- 通过autoencoder来组合不同额外的监督信息, 生成紧凑的监督信号, 对GAN有效.

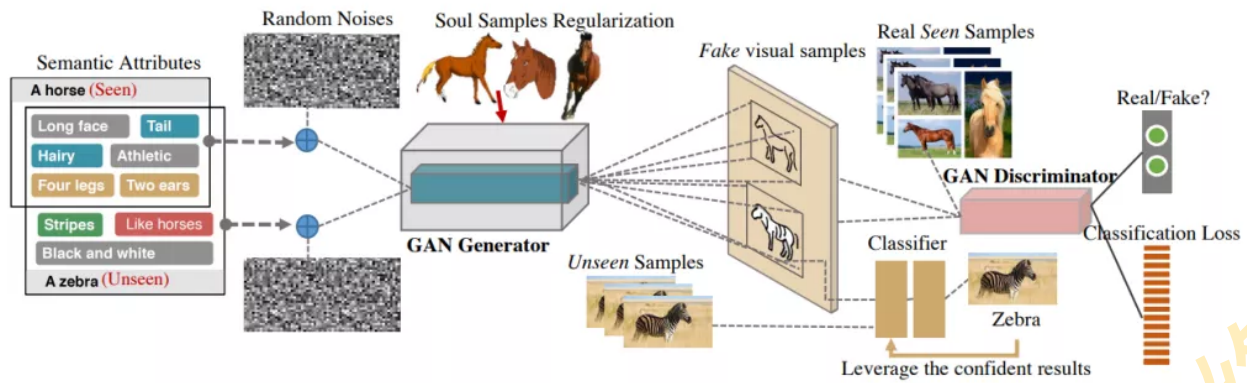
2019_CVPR_LisGAN类别灵魂元向量和对应属性生成视觉向量的表示_GAN训练

- 2019_Leveraging the Invariant Side of Generative Zero-Shot Learning
- http://openaccess.thecvf.com/content_CVPR_2019/papers/Li_Leveraging_the_Invariant_Side_of_Generative_Zero-Shot_Learning_CVPR_2019_paper.pdf

每个类别设置灵魂向量, 通过灵魂向量+对应属性, 可以生成该类别的视觉表示visual.

关键在于:(1) 生成对应类且多样性的visual向量; (2) 确保生成visual向量和真实向量一致.

- 灵魂向量+属性向量+噪声, 生成visual向量:
 - 属性是对应类的属性, 随机噪声受到属性描述的约束
 - 灵魂向量保证是对应的类别. 它必须不是很具体, 可以为每类设置多个灵魂向量(考虑到多视角)
 - 引入损失, 让visual向量尽量只靠近一个灵魂向量
- 确保生成visual和真实向量一致;
 - 用GAN
 - 带鉴别器的鉴别损失,和分类损失.



2019_CVPR_CEWGAN-OD_特征熵分出seen和unseen_GAN生成unseen_重构损失_生成特征类别匹配损失

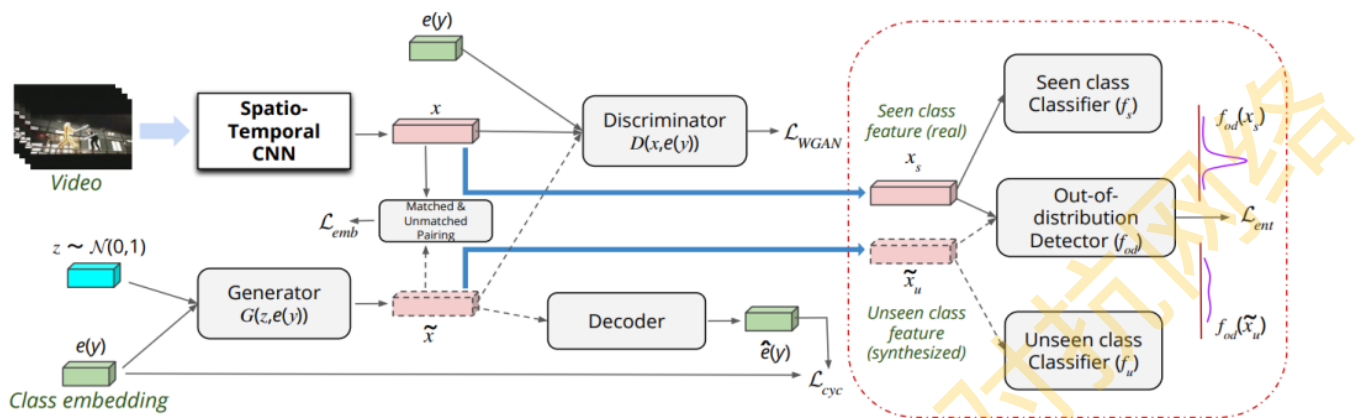
- 2019_Out-of-Distribution Detection for Generalized Zero-Shot Action Recognition
- http://openaccess.thecvf.com/content_CVPR_2019/papers/Mandal_Out-Of-Distribution_Detection_for_Generalized_Zero-Shot_Action_Recognition_CVPR_2019_paper.pdf

在广义Zero-Shot中, 针对seen和unseen进行分别处理. 引入了 out-of-distribution detector 来确定是属于seen还是unseen的动作类别..

对于seen类上进行GAN训练, 从而合成unseen的动作特征. 本文是首次在广义零次学习中提出 out-of distribution detector(OD检测器) 进行视频动作识别.

- seen和unseen的分类:
 - 让OD检测器对seen类生成non-uniform分布(带有peaks), 而对unseen生成uniform分布.
 - 这可以通过最大化unseen动作类别特征的熵来实现.
 - inference时候, OD检测器的输出类别特征的熵, 通过 threshold 即可知道是seen还是unseen.
- GAN生成unseen特征
 - 采用WGAN
 - 在生成视频特征的上下文中, D尝试从合成生成的特征中准确地区分真实视频特征, 而G尝试通过生成语义上与真实特征接近的视频特征来欺骗鉴别器.
 - 语义向量+noise生成 unseen个性化特征
- cycle-consistency loss 重构损失
 - 将GAN生成的unseen特征, Decoder 回原先的语义向量
 - 构建重构损失. 直接用 MSE损失.
- 生成特征的类别匹配损失:

- 对于同一类, GAN生成的特征应该和真实的特征相似
- 不同类时, GAN生成的特征和真实特征应该不相似
- 所以利用cosine相似度, 设置了匹配损失. 让GAN生成的东西具备类别个性.
- 类似TripletLoss



2019_ICCV_通过GAN生成unseen_参数化熵的测度_引导创造力生成

- 2019_Creativity Inspired Zero-Shot Learning
- <https://arxiv.org/pdf/1904.01109.pdf>

基于GAN,通过Wikipedia文本的描述作为条件, 生成unseen类. 通过参数化熵的测度, 从而引导生成unseen类的向量.

2019_ICCV_条件耦合GAN捕获不同任务的联合分布_域自适应 Zeroshot

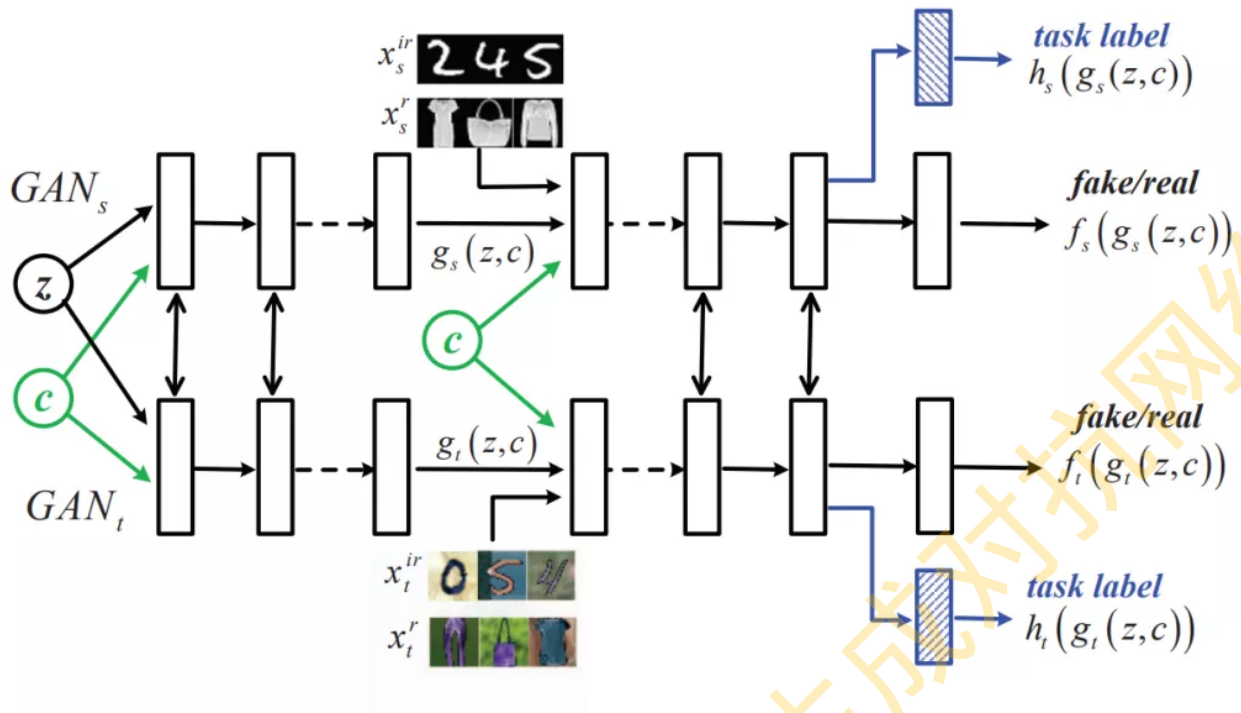
2019_Conditional Coupled Generative Adversarial Networks for Zero-Shot Domain Adaptation

http://openaccess.thecvf.com/content_ICCV_2019/papers/Wang_Conditional_Coupled_Generative_Adversarial_Networks_for_Zero-Shot_Domain_Adaptation_ICCV_2019_paper.pdf

利用耦合GAN 扩展为 条件耦合生成GAN(CoCoGAN) 捕获不同任务 (相关任务RT和无关任务IRT) 中 dual-domain 样本的联合分布.

利用相关任务RT中源样本和IRT中的dual-domain样本来完成域的自适应. 前者提供目标领域的高级概念, 后者提供两个域之间的共享相关性. 文中还提出新的监督信号,让跨任务的表示之间进

行对齐.

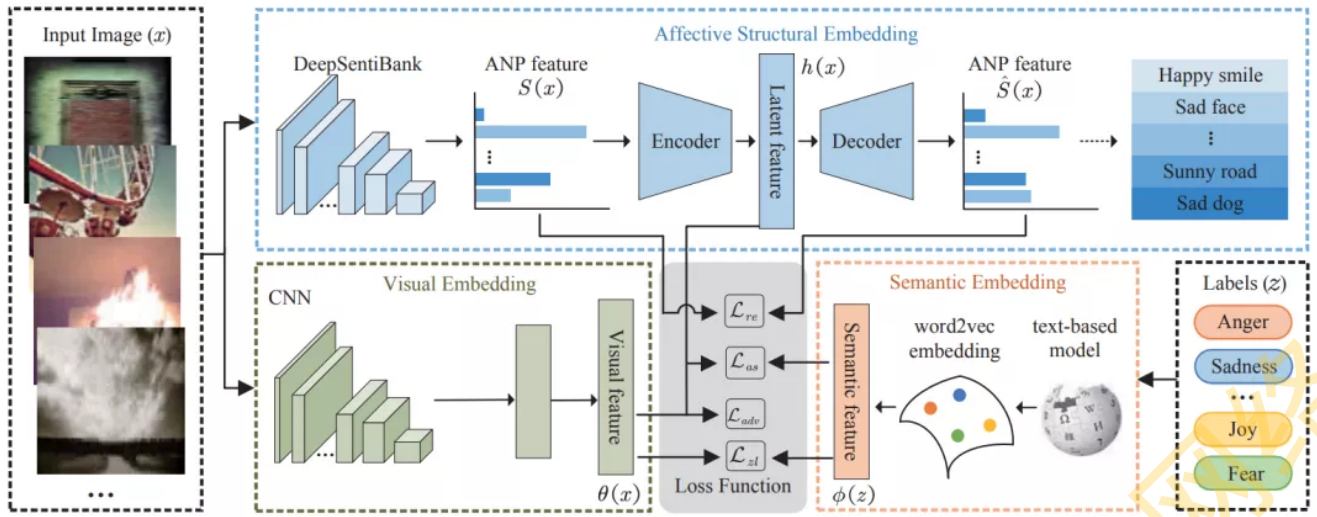


2019_ICCV_情感结构嵌入_首次识别未知情感ZeroShot_中间语义层GAN

- 2019_Zero-Shot Emotion Recognition via Affective Structural Embedding
- http://openaccess.thecvf.com/content_ICCV_2019/papers/Zhan_Zero-Shot_Emotion_Recognition_via_Affective_Structural_Embedding_ICCV_2019_paper.pdf

首次研究识别新出现的未知情感(ZeroShot)

利用中层语义(形容词-名词对)形式构建中间嵌入空间, 从而弥补低级视觉特征和高级语义特征间的gap. 另外加入了对抗约束, 将视觉嵌入和情感嵌入相结合, 让训练中保留视觉特征的辨别力和语义特征的情感结构信息.



公众号近期荐读：

- [GAN整整6年了！是时候要来捋捋了！](#)
- [新手指南综述 | GAN模型太多，不知道选哪个？](#)
- [数百篇GAN论文已下载好！配一份生成对抗网络最新综述！](#)
- [图卷积网络GCN的理解与介绍](#)
- [【CapsulesNet的解析】了解一下胶囊网络？](#)
- [GAN的图像修复：多样化补全](#)
- [CVPR2020之MSG-GAN：简单有效的SOTA](#)
- [CVPR2020之姿势变换GAN：图像里谁都会劈叉？](#)
- [CVPR2020之多码先验GAN：预训练模型如何使用？](#)
- [两幅图像！这样能训练好GAN做图像转换吗？](#)
- [单图训GAN！如何改进SinGAN？](#)
- [有点夸张、有点扭曲！速览这些GAN如何夸张漫画化人脸！](#)
- [见微知细之超分辨率GAN！附70多篇论文下载！](#)
- [天降斯雨，于我却无！GAN用于去雨如何？](#)
- [脸部转正！GAN能否让侧颜杀手、小猪佩奇真容无处遁形？](#)
- [容颜渐失！GAN来预测？](#)
- [强数据所难！SSL（半监督学习）结合GAN如何？](#)
- [弱水三千，只取你标！AL（主动学习）结合GAN如何？](#)
- [异常检测，GAN如何gan？](#)
- [虚拟换衣！速览这几篇最新论文咋做的！](#)
- [脸部妆容迁移！速览几篇用GAN来做的论文](#)

- [【1】GAN在医学图像上的生成，今如何？](#)
 - [01-GAN公式简明原理之铁甲小宝篇](#)
-

GAN&CV交流群，无论小白还是大佬，诚挚邀您加入！

一起讨论交流！长按备注【进群】加入：



更多分享、长按关注本公众号：

