百度飞桨论文复现营

-《First Order Motion Model for Image Animation》论文解读

【论文题目】：First Order Motion Model for Image Animation
【论文作者】：Aliaksandr Siarohin、Stéphane Lathuilière、Sergey Tulyakov、Elisa Ricci、Nicu Sebe
【论文链接】：https://arxiv.org/pdf/2003.00196.pdf
【论文代码】：https://github.com/AliaksandrSiarohin/first-order-model

前言：

7 月 29 日百度飞桨团队正式上线了一门论文复现课程，旨在 28 天内，由百度资深算法工程师和中科院研究员联合授课，手把手带领大家复现 1 篇顶会论文，掌握论文的复现流程。涉及 10 篇论文、2 个领域（GAN 和视频分类），并提供了丰厚的奖励。本篇文章中解读的是 GAN 领域中的其中一篇《First Order Motion Model for Image Animation》，论文解读方法采用吴恩达建议的深读论文流程，简述了该篇论文的创新点、模型结构和最终效果。

课程链接：https://aistudio.baidu.com/aistudio/education/group/info/1340。

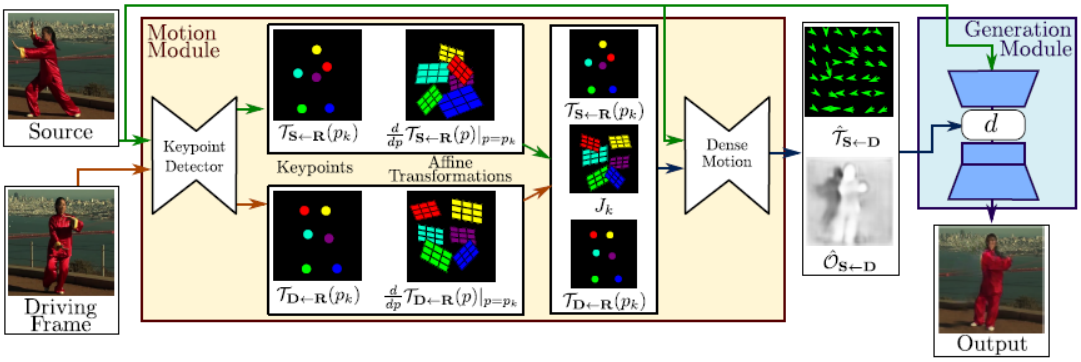
吴恩达教你读论文：https://www.jiqizhixin.com/articles/2020-07-06-3

第一遍：标题（title）、摘要（abstract）和图表（figures）

- 论文标题：图像动画化的一阶运动模型。(本文主要任务实现静态图像到动态图像的转换)
- 摘要：简介实验任务；论文的优势：不需要任何静态图像中对象的标注或先验知识，只需要对某一类对象的动态视频进行训练，之后可以对任意该类对象的静态图像进行动画生成；论文使用方法：通过自监督的方法解耦外观和运动信息，对于复杂动作，通过一组局部仿射变换得到的关键点序列来表示。通过一个生成网络来组合静态图像中提取的外观以及运动视频中的动作；模型跑分效果与开源代码
- 图表：共计 10 张图，1：不同类数据集中模型的生成效果；2：模型结构；3：太极动作数据集中加入模型中不同组件的跑分与结果对比，X2face、Monkey-net 在不同类型数据集中的跑分对比；4：选取两组输入与 X2face、Monkey-net 的结果与跑分对比；5-8：不同类别不同输入数据与 X2face、Monkey-net 的结果对比；9：4 组数据集中的关键点可视化展示；10：不同数据集中动作遮罩结果和生成最终结果的展示。

第二遍：导言（introduction），结论（conclusion），再过一次图表，并简单浏览剩下的内容。

- 导言：简介任务实际应用场景和任务，传统方法使用建立特定对象的模型和计算机图形学技术解决该问题，随着深度学习的发展，GAN 和 VAE 技术被应用于人脸动态图像的面部替换，前期方法存在需要大量环境标注数据集且迁移性不足等缺点。Monkey-Net 通过自监督方法的关键点学习解决了这个问题，但是该方法在关键点邻域中采用了 0 阶模型使得包含大幅度动作变化时生成质量不高。本文在 Monkey-Net 的基础上增加了局部仿射变换来改善复杂动作。其次，提出遮罩生成器，利用上下文信息推断静态图像中不可见的对象部分。使用增加均方差损失的方式优化关键点提取器训练。给出 sota 结果并发布高清太极训练集。
- 结论：通过关键点提取和仿射变换技术设计模型，简述模型的数学描述，通过遮罩推断生成网络中哪一部分图像需要绘制，给出每步计算的 sota 结果。
- 图表及内容浏览：
  此处给出 figure2 的模型结构部分的相关解读：



- Input：静态图片 Source、动作驱动视频 Driving Frame
- Motion module（动作提取模块）：分别对 S（静态图）和 D（动作图）进行关键点检测->提取关键点->仿射变换，结合 S 输入提取密集动作输出关键点到 S 的映射关系以及遮罩（用于推断需要通过绘制得到的部分）
- Generation Module（图像生成模块）：通过提取静态图像的特征信息融合动作提取模块的结果最终生成输出结果

第三遍：阅读论文的整个部分，但跳过任何可能对你来说陌生的复杂的数学或技术公式

除去文章的引言和结论，该篇论文还有三大部分的内容，releated work、method 和 experiments。

- releated work: 简述视频生成中的技术提及 VAE、MOCOGAN 等，简述了文中借鉴的主要思想，在 figure2 中已介绍；简述图像动画化，主要指人脸、姿势识别等领域，提及 img2img，加强 GAN，X2Face 和 MonkeyNet 等研究。介绍了本文算法与这些提及算法之间的联系
- Method：给出了详细的算法结构及其数学推导依据，重点讲述了局部仿射变换、遮罩图像生成以及训练的损失函数。并给出了关联动作变换的测试结果
- Experients：最后给出了一些训练经验和技巧，从数据集、评估方法、一些指标的定义、与 X2face 和 MonkeyNet 之间的对比研究和 SOTA 的对比。

以上为该篇论文的论文解读，因为时间有限，许多问题没有展开叙述。之后会给出实现代码的详细解读，届时会对照代码介绍论文的核心方法，被给出 paddlepaddle 的最终复现结果。感兴趣的同学也可以参加一波论文复现课程，欢迎大家批评指正~