

# week13 论文笔记

---

## week13 论文笔记

[A Survey on Vision Transformer](#)

[Deep-Emotion Facial Expression Recognition Using](#)

[Robust Facial Expression Recognition with Convolutional Visual Transformer](#)

[TransFER- Learning Relation-aware Facial Expression Representations with Transformers](#)

这周主要看的是Vision Transformer研究transformer的介绍论文，包括一些综述和这个方面的最新技术进展

## A Survey on Vision Transformer

---

### 论文摘要

transformer首先应用于自然语言处理领域，是一种以self-attention机制为基础的深度神经网络。由于其强大的表现能力，思考将其用于计算机视觉任务。在各种视觉任务测试中，基于transformer的模型的性能与其他类型的模型具有相似甚至更好的效果，因而transformer正受到计算机视觉越来越多的关注。本篇论文就是对现有的计算机vision transformer进行介绍，然后将它们划分到不同的视觉任务中，并分析其优缺点，最后为其发展提供一些方向与建议。

### 解决的问题

对现有的计算机视觉领域的vision transformer技术进行了整理

### 做的工作

1. 简要介绍了transformer的self-attention机制以及其如何迁移至计算机视觉领域
2. 详细介绍vision transformer模型：
  - 骨干网络
  - 高/中/低水平视觉
  - 视频任务
3. 近年来的一些高效的vision transformer技术
  - 并列了一张表格整理了它们的视觉任务类型、特点或创新点，发布时间与平台
  - 并比较了它们在各种计算机视觉任务中与传统方法进行比较之后的结果
4. 总结了当下的几个研究方向以及未来的前景

### 结论

在视觉领域，从前CNN深度神经网络非常热门，而近几年，越来越多人开始关注vision transformer，因为transformer具有高性能和低消耗的特点，所以之后一段时间计算机视觉领域的目标将会变成开发高效有效的vision transformer

# Deep-Emotion Facial Expression Recognition Using

## 论文摘要

提出了一种基于注意卷积网络的深度学习方法，能够将注意力集中在重要的部分上，从而帮助计算机视觉的识别任务，并在各个人脸表情数据集上有着很好的效果

## 解决的问题

传统的方法是依赖于手工特征，且在一些受控条件下的实验室数据集上能够表现良好，但在更具有挑战性的其他数据集或其他视觉任务中不能很好的执行

## 做的工作

1. 提供了相关的一些技术工作的介绍与概述
2. 详细的介绍了论文所提出的基于注意力的卷积网络的模型框架与体系结构并对其进行解释
3. 提供了模型在包括FER-2013, CK+, FERG, JAFFE等各个数据集的实验结果，并将实验结果可视化
4. 对模型和工作进行总结

## 结论

对于特殊区域的特征提取在人脸识别工作中是非常重要的，且发现对于不同情绪在不同特殊区域是有不同比重的。这种方法能够使模型具有更强的竞争力

# Robust Facial Expression Recognition with Convolutional Visual Transformer

## 论文摘要

提出了一种新模型CVT以适用于野外表情数据集，这种模型利用两种特征的注意选择性融合(ASF)由两个分支的CNN生成的特征图，融合后的特征图被压平并映射到视觉单词序列中，最终进行识别。并对这个模型在野外数据集上进行了效果评估

## 解决的问题

面部表情识别(FER)因为遮挡，不同的头姿势，脸在无约束条件下的变形和运动模糊情况而导致的识别困难。识别任务在野外是极具挑战性的

## 做的工作

1. 简要回顾了FER在野外的相关工作并提供了FER的最新进展
2. 提出了CVT模型，并对此模型的核心思想和框架体系进行了介绍
3. 在三个公开的数据集RAF-DB、FERPlus和AffectNet上进行了实验和评估
4. 对工作进行了总结

## 结论

1. 提出了CVT模型，集成了LBP特征和CNN特征，即考虑了全局和局部特征，使得人脸表情识别精度提高
2. 设计了一种简单有效的特征融合模块ASF来聚合全局和局部的面部信息

3. 在数据集上的实验结果表明，该算法具有很好的泛化能力

## TransFER- Learning Relation-aware Facial Expression Representations with Transformers

### 论文摘要

提出了一种学习面部局部区域之间关系的模型TransFER，它具有三个部分：MAD、ViT-FER、MSAD。并将其和当前一些最先进的方法比较，显示了它的有效性和实用性

### 解决的问题

当前工作对于局部面部区域的特征以及它们之间的关系关注不够

### 做的工作

1. 介绍了相关的工作
2. 设计了MAD模块和MSAD模块，不仅关注来自不同局部区域的特征，而且为了防止多个自我关注可能会取到类似的特征，还利用随机去除自注意力模块，使模型能够学习到不同局部之间的丰富关系
3. 详细描述了TransFER的框架体系
4. 在几个具有挑战性的数据集上进行实验
5. 总结工作

### 结论

模型TransFER可以学习丰富、多样的关系感知的局部特征，并且超过了当下最先进的方法