# 阅读报告

English 中文

## 一、论文标题：

Taming Transformers for High-Resolution Image Synthesis

## 二、作者单位、发表时间：

2020-2021

## 三、论文背景和问题：

前diffusion时代，Transformer关注全局但计算成本高昂，CNN可以有效处理局部特征，但难以捕捉全局信息。作者综合了两者，提出了VQ-GAN和transformer结合生成高清图像的新方法。

## 四、论文动机和贡献解读：

## 五、方案设计详细分析：

VQ-GAN基于VQ-VAE做了两点改进：损失函数改为感知误差（perceptual loss），增加了基于图块的判别器（Patch GAN），得到一个强大的图像压缩模型。

之后，利用transformer生成压缩图像。由于transformer计算开销大，难以生成大图像，作者采取了滑动窗口机制，将图片分成16x16的图块，每次只对16x16个图块进行计算。

## 六、实验效果及其分析：

## 七、结论：

## 八：思考