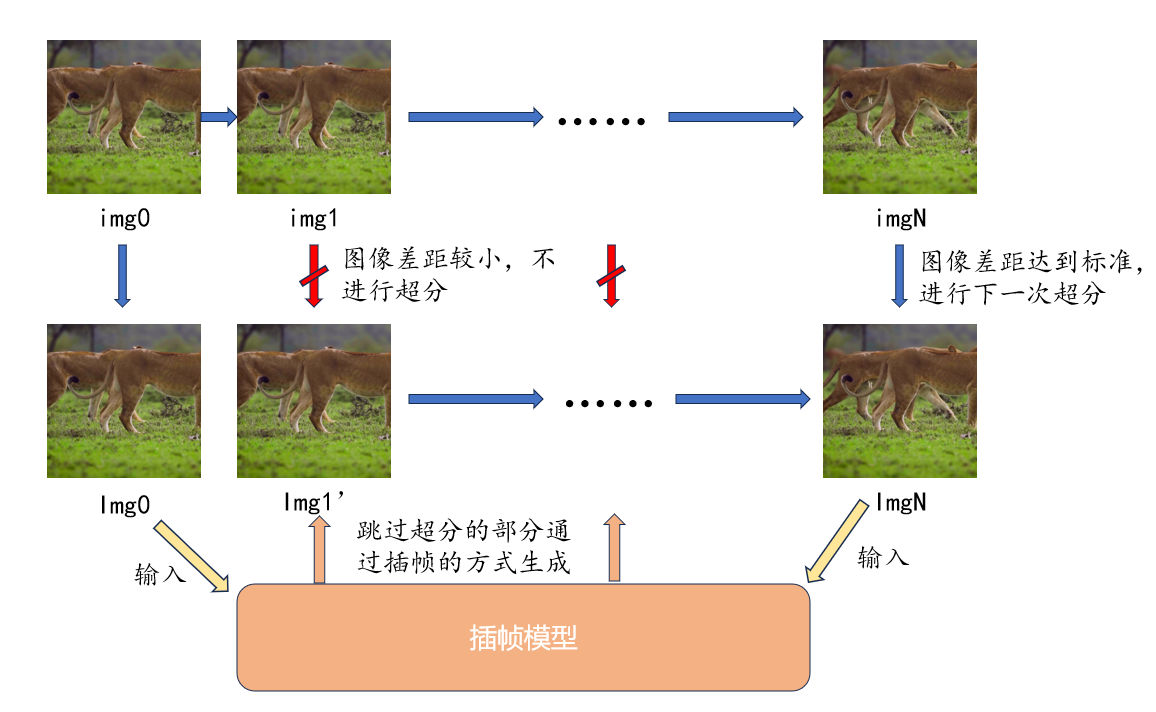
**插帧算法和帧间相似度判断**

1. 示意图和详细描述



由于在实际应用中，视频的帧之间相似度极高，如果对每帧都进行超分，会造成大量的计算量冗余。因此，本方案尝试对单张图像超分后，逐帧向后对比，直到出现差异较大的图像再进行下一步的超分。得到相邻的两张超分输出之后，再使用插帧模型生成两帧之间跳过超分的部分，进而得到最终的完整视频。 这种通过插帧方式得到的视频，还可以避免diffusion不稳定性造成的前后帧不匹配，不容易出现闪烁，视频更佳平滑。

其中，帧间相似度可以使用余弦相似度进行计算，即：

其中，每个像素点都有RGB三通道维度，也就对应了一个三维空间中的向量，两个向量点乘等于其模相乘再乘以夹角的余弦值，因此可以用来计算图像的相似度。类似地，也可以使用均方误差（MSE），结构相似性（SSIM），峰值信噪比（PSNR）等多种方式进行计算。

1. 实验设计和加速效果讨论

初步实验使用了现有的RIFE插帧模型，其来自一篇发表在2022ECCV上的文章。

通过初步实验，尝试了其在512x512分辨率上的插帧效果如下表所示：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **插帧实验** | | | | | |
|  | 1：1 | 1：3 | 1：7 | 1：15 | 对照 |
| SSIM↑ | 0.998 | 0.994 | 0.946 | 0.885 | 1 |
| psnr↑ | 67.998 | 56.661 | 43.7 | 35.355 | 80 |
| lpips↓ | 0.002 | 0.005 | 0.023 | 0.055 | 0 |
| musiq↑ | 67.429 | 67.227 | 66.863 | 66.599 | 66.97 |
| clipiqa↑ | 0.541 | 0.559 | 0.557 | 0.54 | 0.501 |
| niqe↓ | 3.986 | 4.064 | 4.391 | 4.467 | 4.023 |

其中，组别的1：1代表了原始图片和插帧生成图片的数量相等，而1：15代表了插帧生成图像为原始图像数量的15倍。表中的对照组由GT直接测试得到。由实验数据可以看出，插帧的密度越高，其各指标逐步变差。但下降的幅度在1：3和1：7实验组都较为轻微。

而时间上，在V100上，一个53S，25帧，1280x720的视频，增加到50帧，所花费的时间为63s，平均单帧耗时47.5ms。单张的512x512图像插帧的平均耗时约10ms，此速度显著快于超分，因此可以实现提升整体速度的目的。

而为了进一步验证这一方案能够减少整体时间开销并不影响视频质量，后续还会结合超分模型，将全视频帧超分得到的结果与使用本方案得到的结果进行图像相似度，图像质量，视频一致性以及时间开销的对比。