



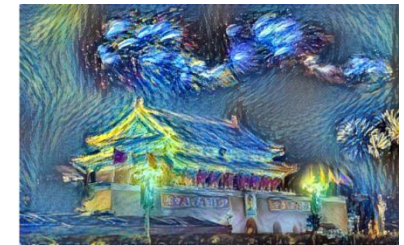
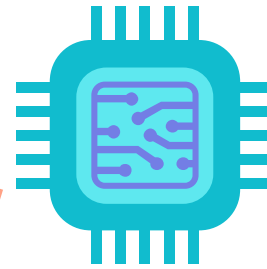
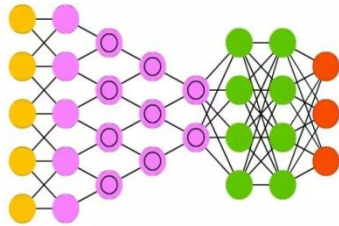
智能计算系统

实验二 基于TF风格迁移

算法的实现

中国科学院计算技术研究所

Driving Example



神经网络



深度学习



编程框架



Bang



平台架构



实验一



实验二

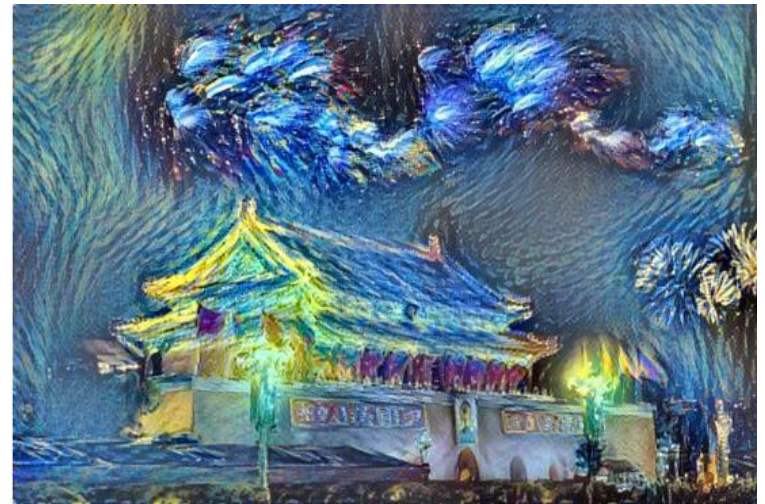


实验二将基于寒武纪系列开发平台上完成Tensorflow风格迁移算法的实现

Driving Example



像艺术家一样



提纲

- ▶ 实验目的与要求
- ▶ 背景介绍
- ▶ 实验步骤
- ▶ 演示运行

实验目的

- ▶ 掌握使用TensorFlow定义完整网络结构的方法
- ▶ 掌握使用TensorFlow进行神经网络预测的方法
- ▶ 理解模型量化的基本原理和方法，具备使用TensorFlow对模型进行量化的能力
- ▶ 掌握在DLP上使用TensorFlow对模型进行预测的原理和方法

实验要求

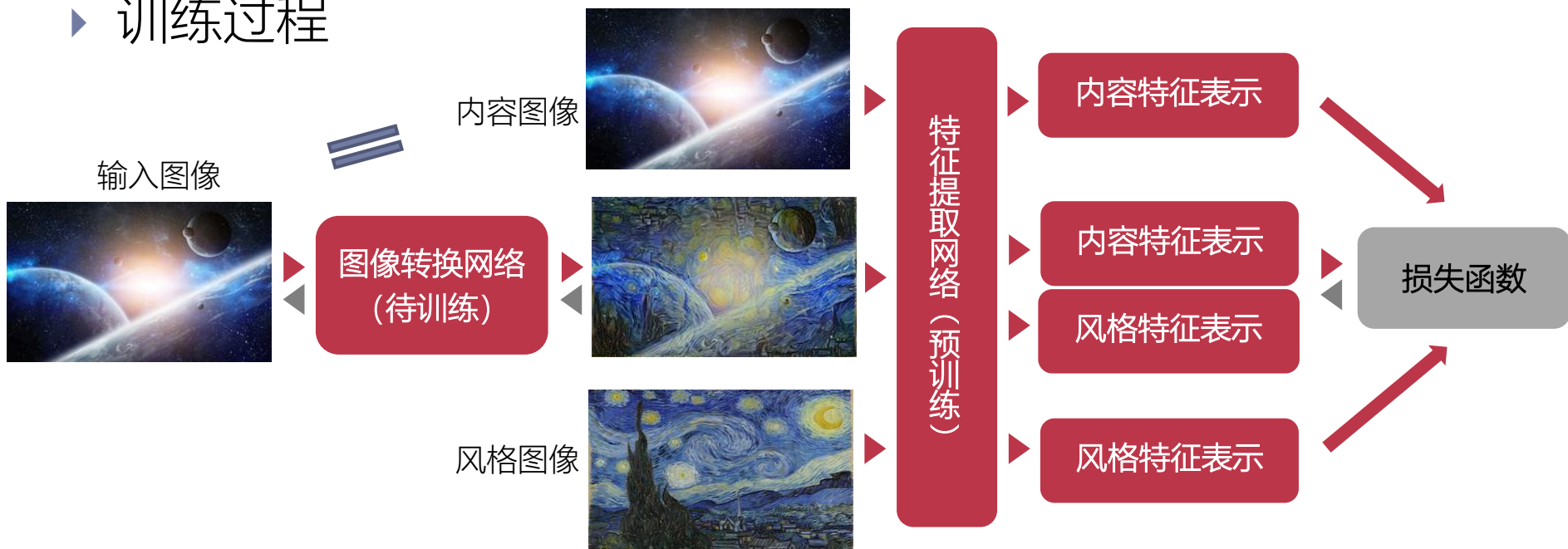
- ▶ 必做题1：基于TensorFlow的实时风格迁移预测
- ▶ 必做题2：以下算法3选1
 - ▶ 目标检测-YOLOv3 (难度系数分1.1)
 - ▶ 文本识别OCR-EAST (难度系数分1.0)
 - ▶ 自然语言处理-BERT (难度系数分1.2)

提纲

- ▶ 实验目的与要求
- ▶ 背景介绍
- ▶ 实验步骤
- ▶ 演示运行

实时图像风格迁移

▶ 训练过程

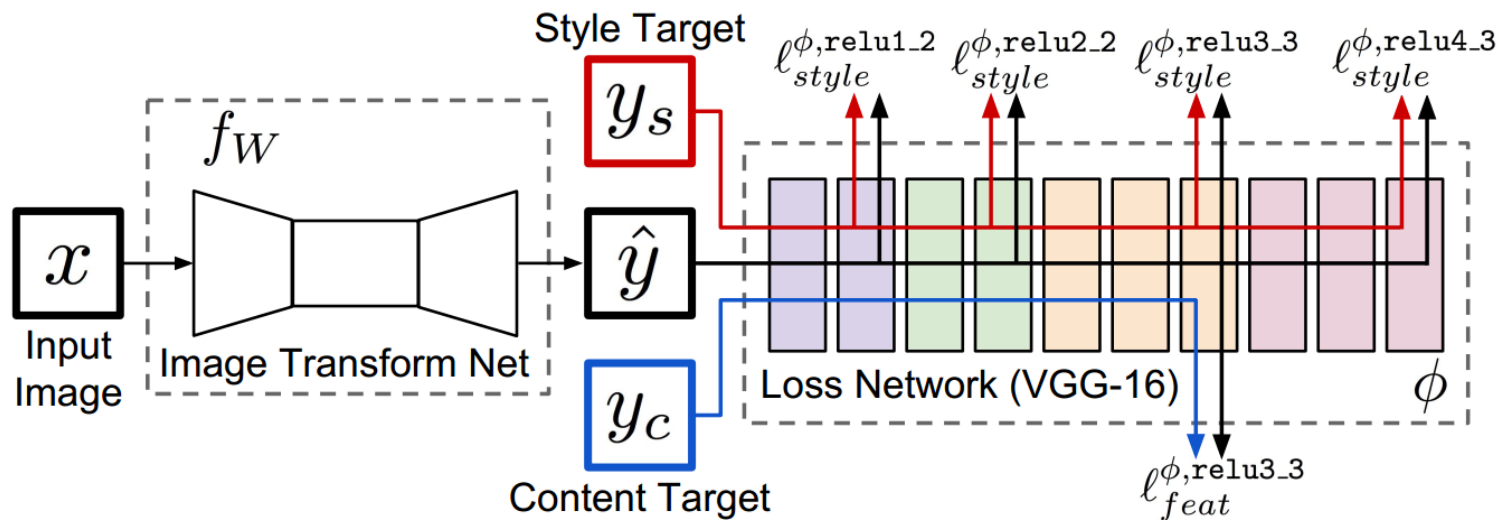


▶ 实时转换过程



实时图像风格迁移

- ▶ Johnson J, Alahi A, Feifei L, et al. Perceptual Losses for Real-Time Style Transfer and Super-Resolution[J]. 2016.
- ▶ 项目实现: <https://github.com/jcjohnson/fast-neural-style>



- ▶ Image Transform Net
 - ▶ 深度卷积网络, 参考DCGAN的设计思想: 用步幅卷积替代pooling、每个卷积层后接BatchNorm和Relu;
 - ▶ 增加残差结构, 使网络更易训练;

实时图像风格迁移



Content

Gatys et al.

Ours

Image Size	Gatys <i>et al.</i> [11]			Ours	Speedup		
	100	300	500		100	300	500
256 × 256	3.17	9.52s	15.86s	0.015s	212x	636x	1060x
512 × 512	10.97	32.91s	54.85s	0.05s	205x	615x	1026x
1024 × 1024	42.89	128.66s	214.44s	0.21s	208x	625x	1042x

模型量化

- ▶ DLP硬件支持定点数据类型运算，对mlp、conv等算子可使用int数据类型来提高计算效率
- ▶ Int8类型数值*i*表示的32位浮点数*value*为：

$$value = \frac{i \times 2^{position}}{scale}, \quad i \in [-128, 127]$$

- ▶ Position为指数因子，scale为缩放因子，
- ▶ 反向推导上述公式，可得：

$$i = \frac{value \times scale}{2^{position}}$$

深度学习推理

- ▶ 基于DLP实现深度学习推理有两种方式：
 - ▶ 在线推理：通过TensorFlow调用CNML库，将模型文件在线编译生成融合的DLP指令，再调用CNRT执行计算得到输出
 - ▶ 离线推理：预先对TensorFlow生成的网络模型文件进行编译，生成包含DLP指令集的模式文件，使用时直接使用CNRT加载模型实施推理

提纲

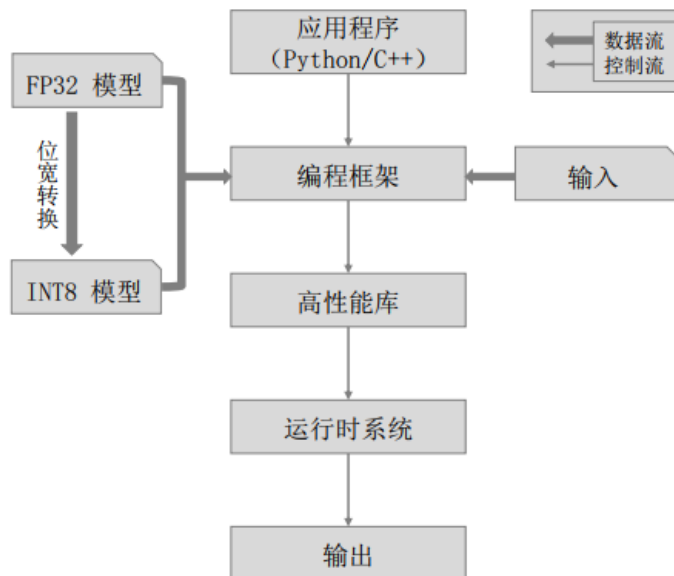
- ▶ 实验目的与要求
- ▶ 背景介绍
- ▶ 实验步骤
- ▶ 演示运行

1、模型量化

- ▶ 将原始的风格迁移算法中，float32数据类型的pb模型量化成int模型
- ▶ 执行python fppb_to_intpb.py udnie_int8.ini
- ▶ 生成的模型保存在：
 - ▶ opt/AICSE-demo-student/demo/style_transfer_bcl/models/int_pb_models

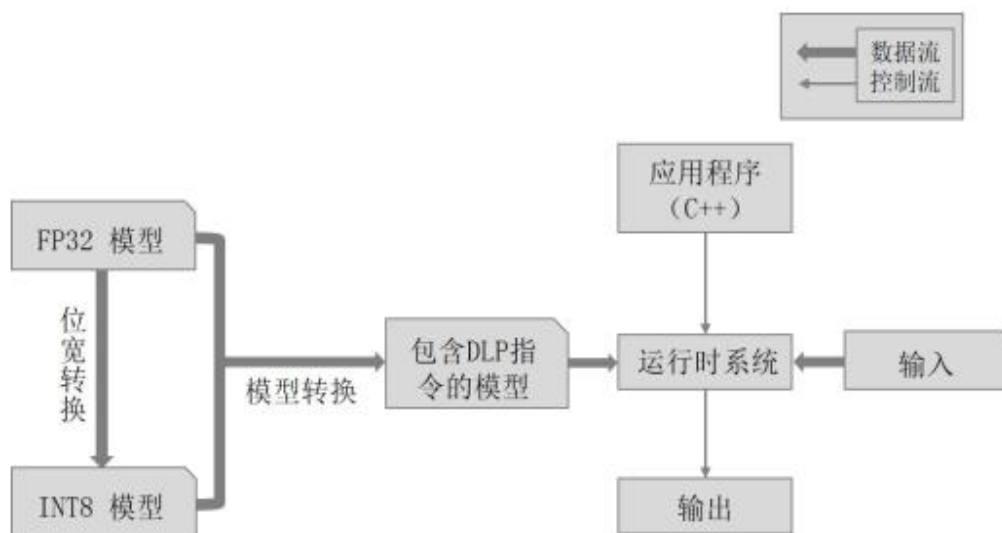
2、在线推理

- ▶ 分别使用dlp和cpu来实现实时风格迁移的在线推理程序，比较两种平台实现的精度差异
- ▶ 通过配置文件生成并保存两种离线模型（原始模型、PowerDifference算子模型）



3、离线推理

- ▶ 将在线推理过程中生成的离线模型保存至特定文件夹下，
补全CNRT离线推理模型代码，编译并执行离线推理



提纲

- ▶ 实验目的与要求
- ▶ 背景介绍
- ▶ 实验步骤
- ▶ 演示运行

演示运行



谢谢大家!