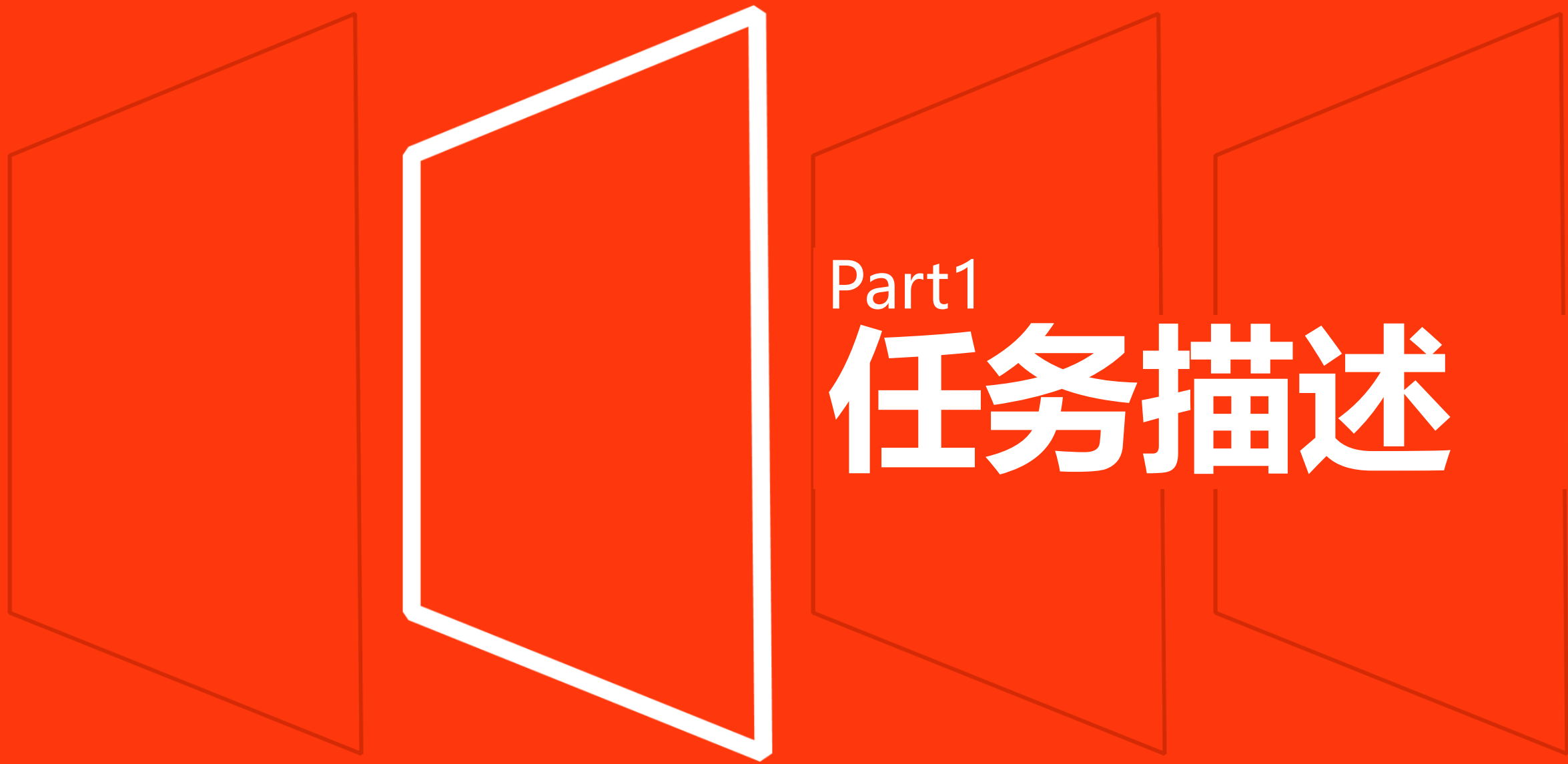
PRESENTED BY 张舸

猫狗图像识别

迁移学习实现猫狗分类

CONTENT 目录


Part1
任务描述

Part One 任务描述

对于我们人来说，在一张图片中区分出猫与狗是件再容易不过的事，但是我们想要利用计算机对猫与狗做区分，需要做什么，如何做？计算机不像我们人类能先用眼睛观察，然后进行区分，而计算机能做的是对数据进行处理与传输，所以我们的任务也就出来了，如何去挖掘图像背后的有用数据，找到规律，怎么建立数学模型，怎么训练模型，怎样验证模型，如何使得计算机区分猫与狗的能力最大、误差最小。即最后通过输入一张图片预测这张图片当中的动物是猫还是狗，可以通过概率进行表达。

Part One 挖掘数据集

数据集

在本任务当中选取两个不同的数据集：训练集、验证集；训练集包含25000张图片，其中猫狗比例大致为50%；验证集为12500张图片其中有百分之二十来自训练集；从中选取了几张图片，如右图所示。







Part2

问题定义

Part Two 问题定义

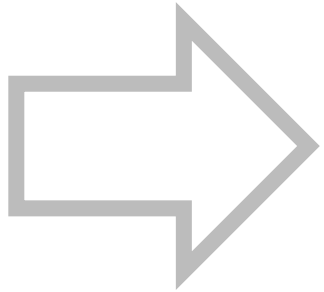
问题

首先我们需要对图片进行怎么的处理以便对图片进行数据的提取与预处理，然后是建立怎样的数学模型合适。


Part 3
解决方案

主要问题

- 数据量较大
  - 训练集：25000张×N，545MB×N
  - 测试集：12500张，272MB
- 计算资源不足
  - GPU：1050M
  - CNN模型：3个，＞100层
  - 计算一次前向传播：＞20min



迁移学习

整体思路

- 图像预处理
- CNN提取特征向量
  - 低级特征：直线，圆弧，颜色
  - 更高级的特征：鼻子，眼睛，
- 对特征向量进行二分类
  - 传统机器学习：SVM，随机森
  - 深度学习：CNN，FNN

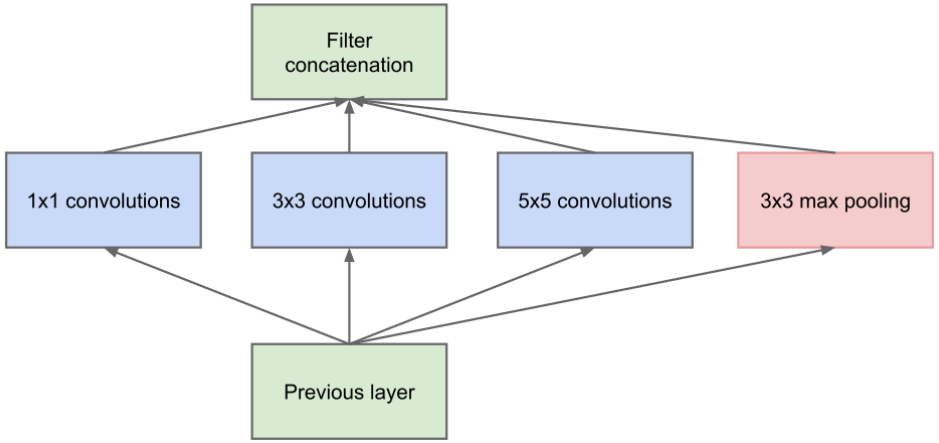
等效于固定CNN前半部分网络的参数，只训练后半部分网络

训练时，无论迭代次数为何，前半部分网络对相同的图像只需经过一次前向传播的计算，大大减少了训练时间

1epoch: 30min
10epochs: 300min → 30min

Inception系列网络

——” We need to go deeper.”

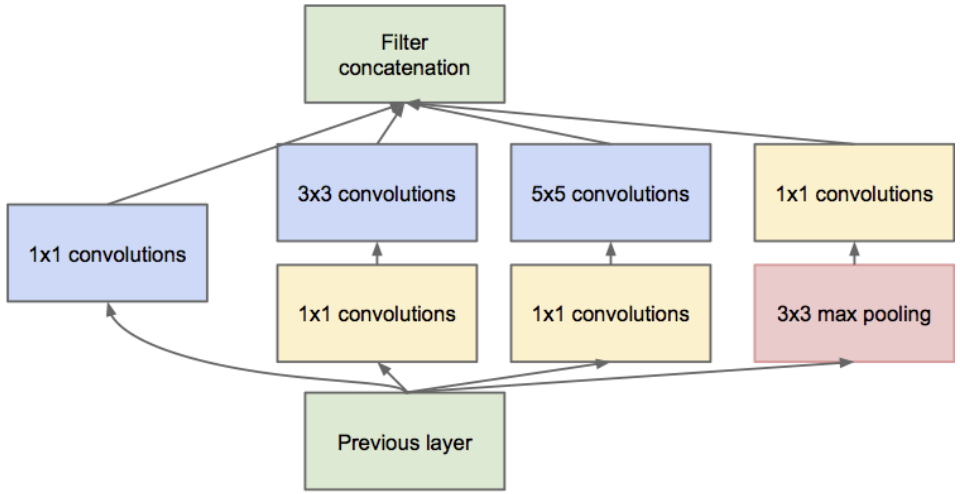


- Inception
  - 让网络学习不同卷积核提取出的特征

- Xception

Inception系列网络

——” We need to go deeper.”



- Inception
  - 让网络学习不同卷积核提取出的特征
  - 引入1×1卷积核

- Xception

Inception系列网络

——” We need to go deeper.”

- Inception
  - 让网络学习不同卷积核提取出的特征
  - 引入1×1卷积核
  - 两个3×3卷积核代替5×5卷积核

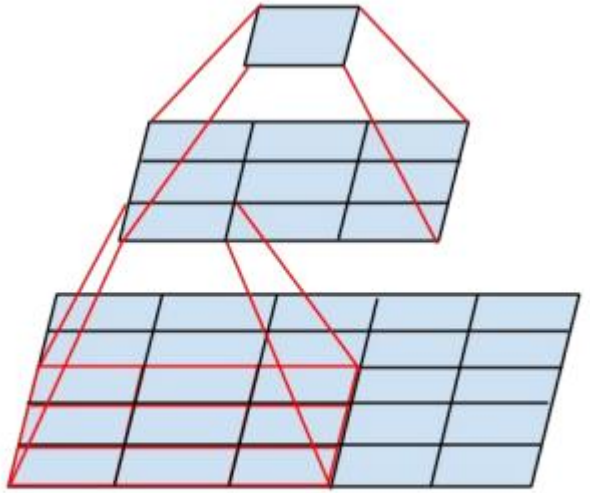


Figure 1. Mini-network replacing the 5×5 convolutions.

- Xception

Inception系列网络

——” We need to go deeper.”

- Inception
  - 让网络学习不同卷积核提取出的特征
  - 引入1×1卷积核
  - 两个3×3卷积核代替5×5卷积核
  - 引入1×n和n×1的非对称卷积核

- Xception

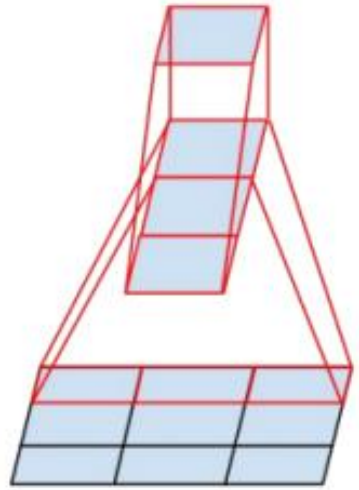


Figure 3. Mini-network replacing the 3 × 3 convolutions. The lower layer of this network consists of a 3 × 1 convolution with 3 output units.

Inception系列网络

——” We need to go deeper.”

- Inception
  - 让网络学习不同卷积核提取出的特征
  - 引入1×1卷积核
  - 两个3×3卷积核代替5×5卷积核
  - 引入1×n和n×1的非对称卷积核

- Xception
  - Extreme Inception: 完全分离通道间的相关性和空间上的相关性

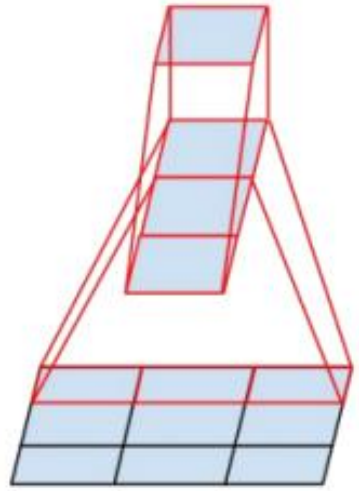
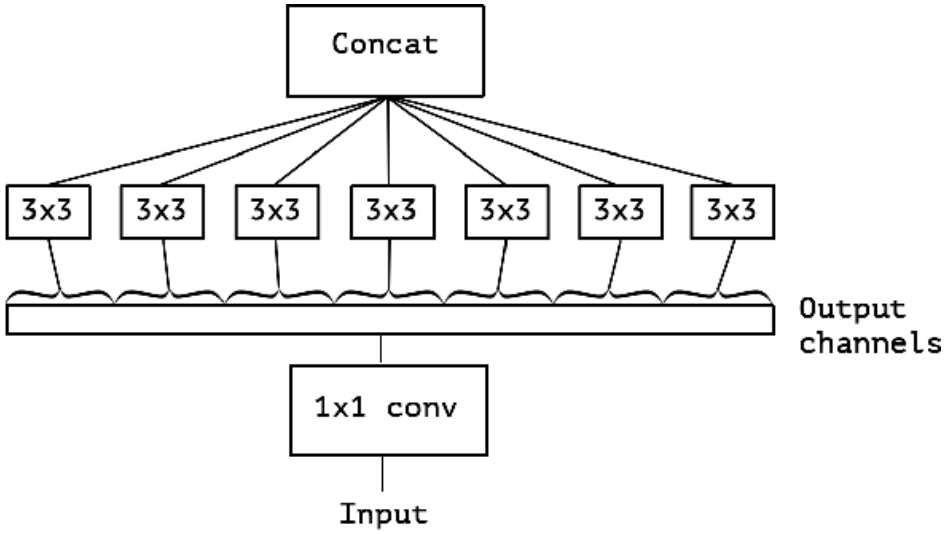
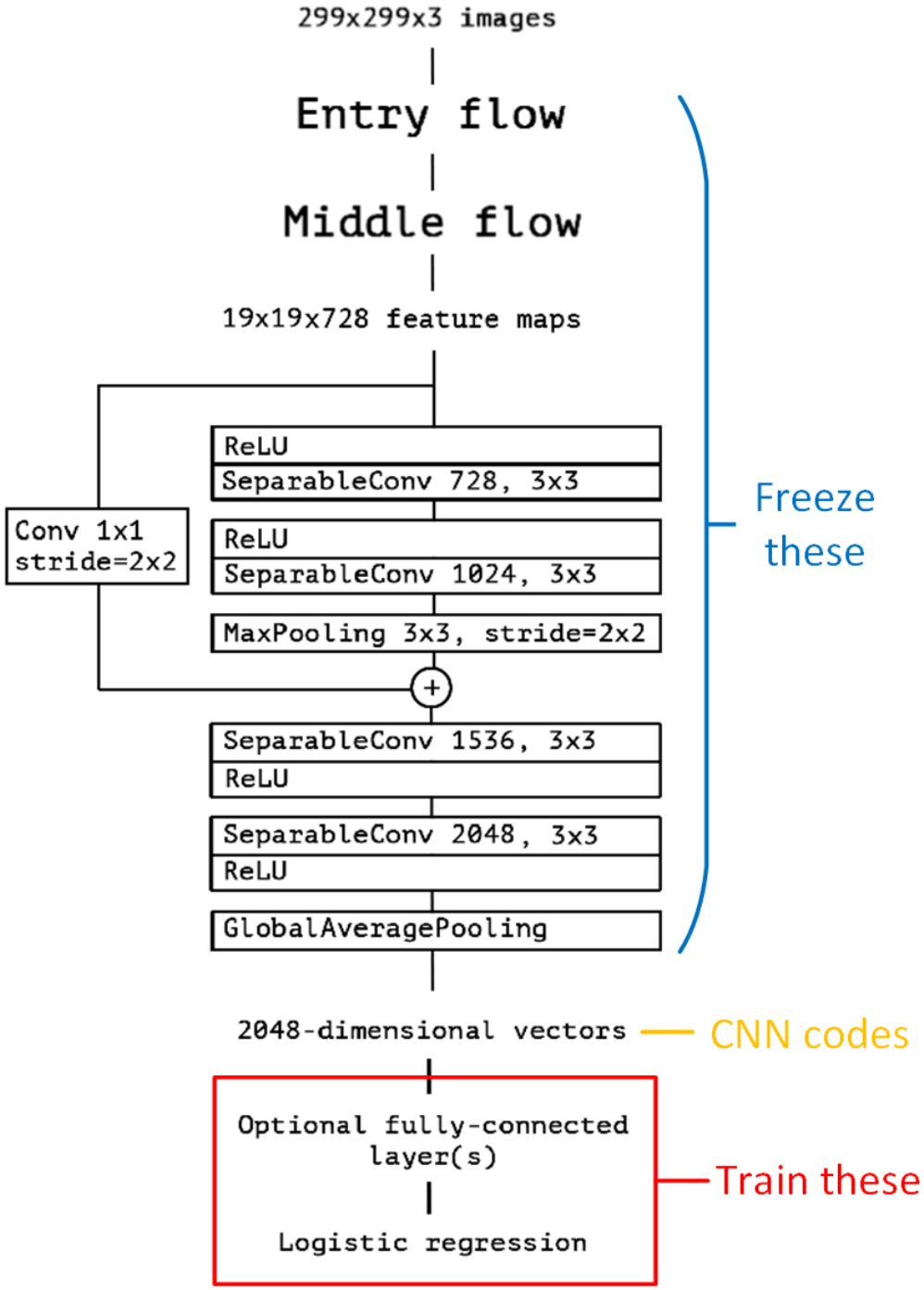


Figure 3. Mini-network replacing the 3 × 3 convolutions. The lower layer of this network consists of a 3 × 1 convolution with 3 output units.



迁移学习

- 三种主要方案
  - ConvNet as fixed feature extractor
  - Fine-tuning the ConvNet
  - Pretrained models
- 特征提取
  - 保留较强的泛化能力，同时最大程度减少计算量

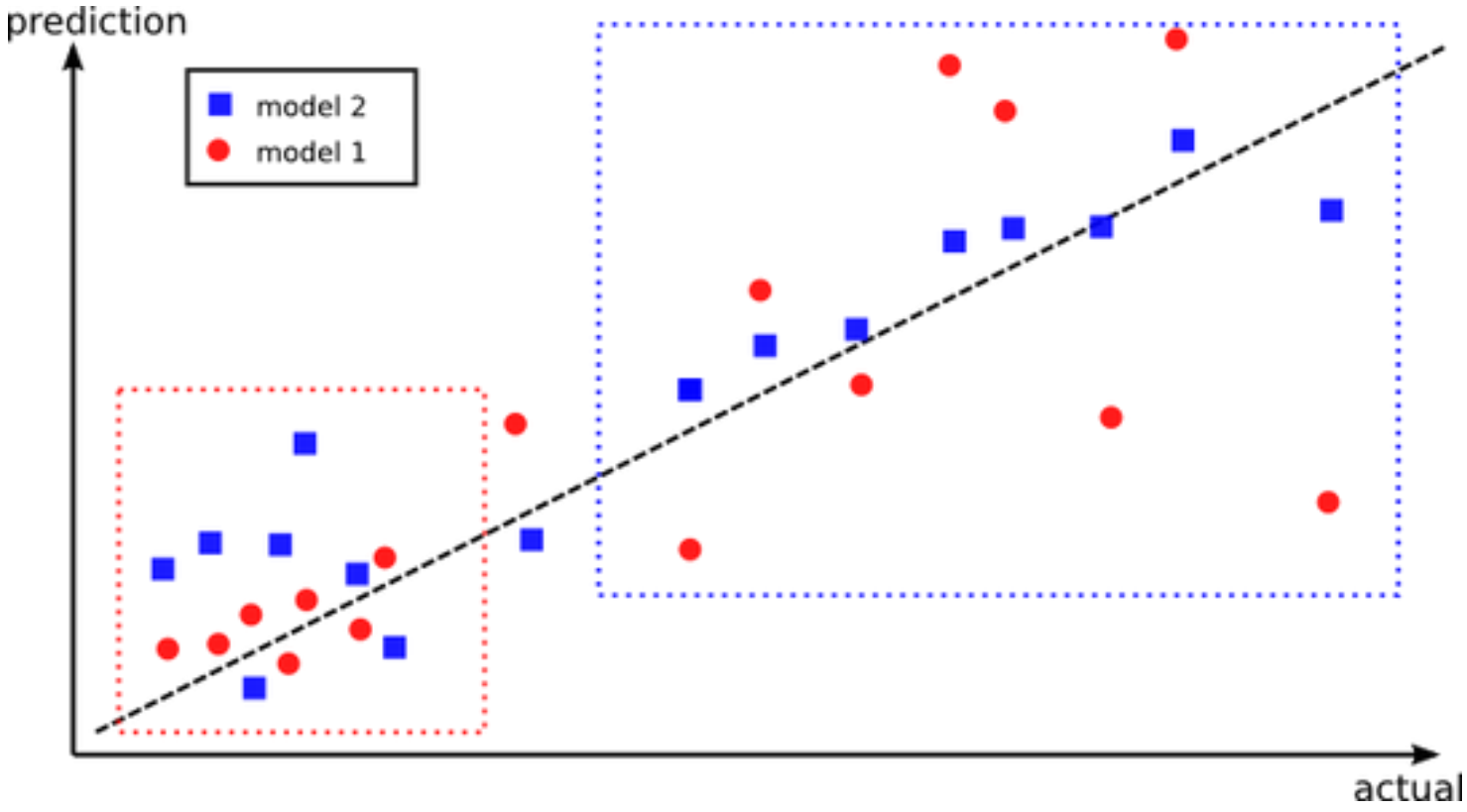


特征向量处理

- SVM (baseline model)
  - 减少计算量：增加全连接层，(2048) → (20)
- FNN
  - 防止过拟合：dropout

模型融合

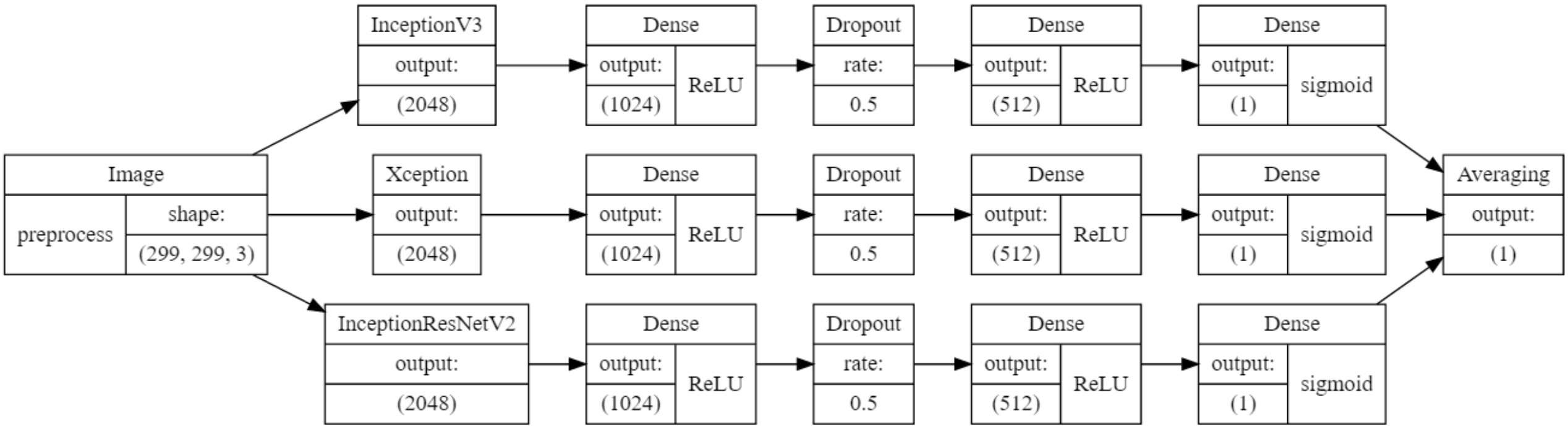
- 方法
  - Voting
  - Averaging
  - Bagging
  - ……
- 作用
  - 降低误差和方差，防止过拟合



备选方案

- 单CNN模型+SVM/FNN分类
- 多CNN模型+FNN分类
  - 扩展了特征向量 → 更全面的特征信息
- 多CNN模型+FNN分类+模型融合
- ……

Part Three 解决方案




Part 4
数据处理

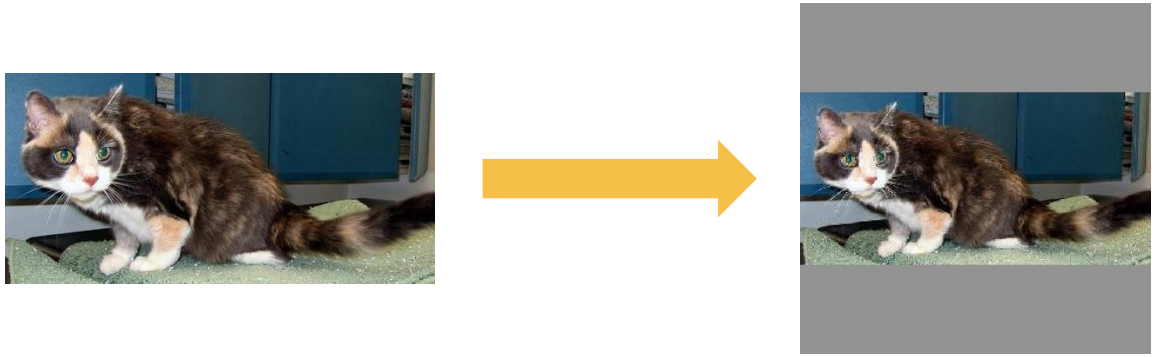
Part Four 数据处理步骤及结果

- 将训练集按标签分类

  由于使用的keras框架需要将不同种类的图片分在不同的文件夹中，因此把训练集的25000张图片按标签分为猫狗两类，置于生成的“cat”和“dog”两个子文件夹，每类各12500张图片。

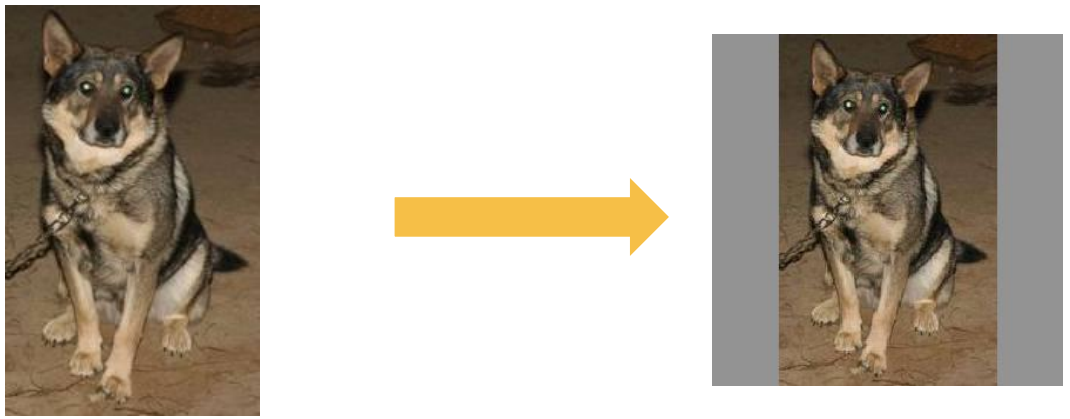
- 统一图片尺寸

  由于原始图片大小不一，为便于处理，先将其置于以长边为边长的背景正方形中央，再等比例放缩成299×299大小的图片，如右图：



- 目前尝试的图像处理方法及所得效果

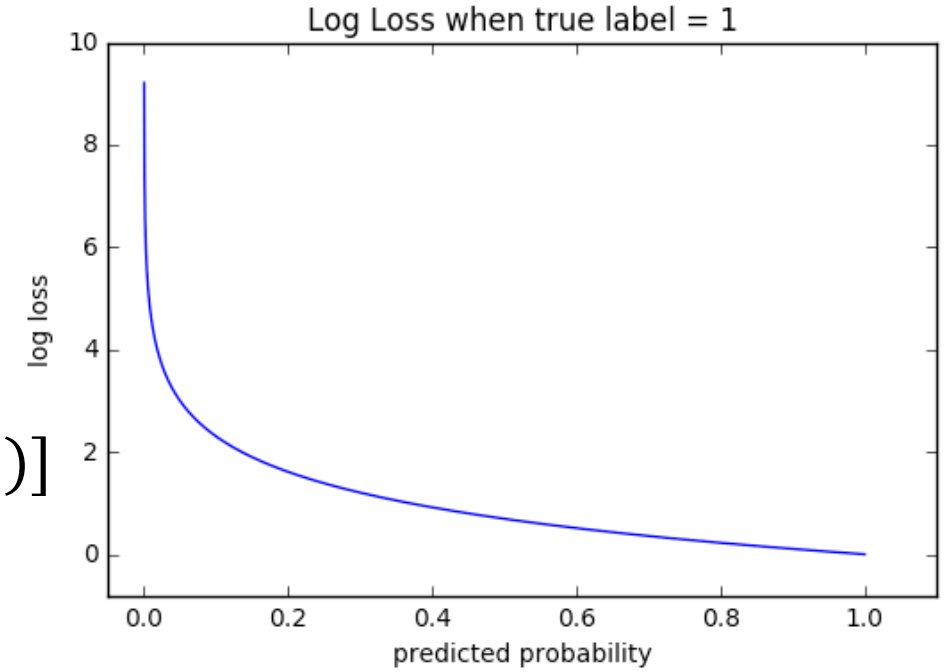
  ①水平翻转：增加了一些验证集的准确率但对测试集loss的影响不大；
  ②椒盐噪声：准确率反而下降了；
  ③直方图均衡：暂没看出影响。




Part 5
结果分析

单模型+SVM/FNN

- Log Loss: 0.05709 / 0.04475
  - $-\frac{1}{n}\sum_{i=1}^{n}[y_i \log(\hat{y}_i) + (1-\hat{y}_i)\log(1-\hat{y}_i)]$
  - 102nd / 28th

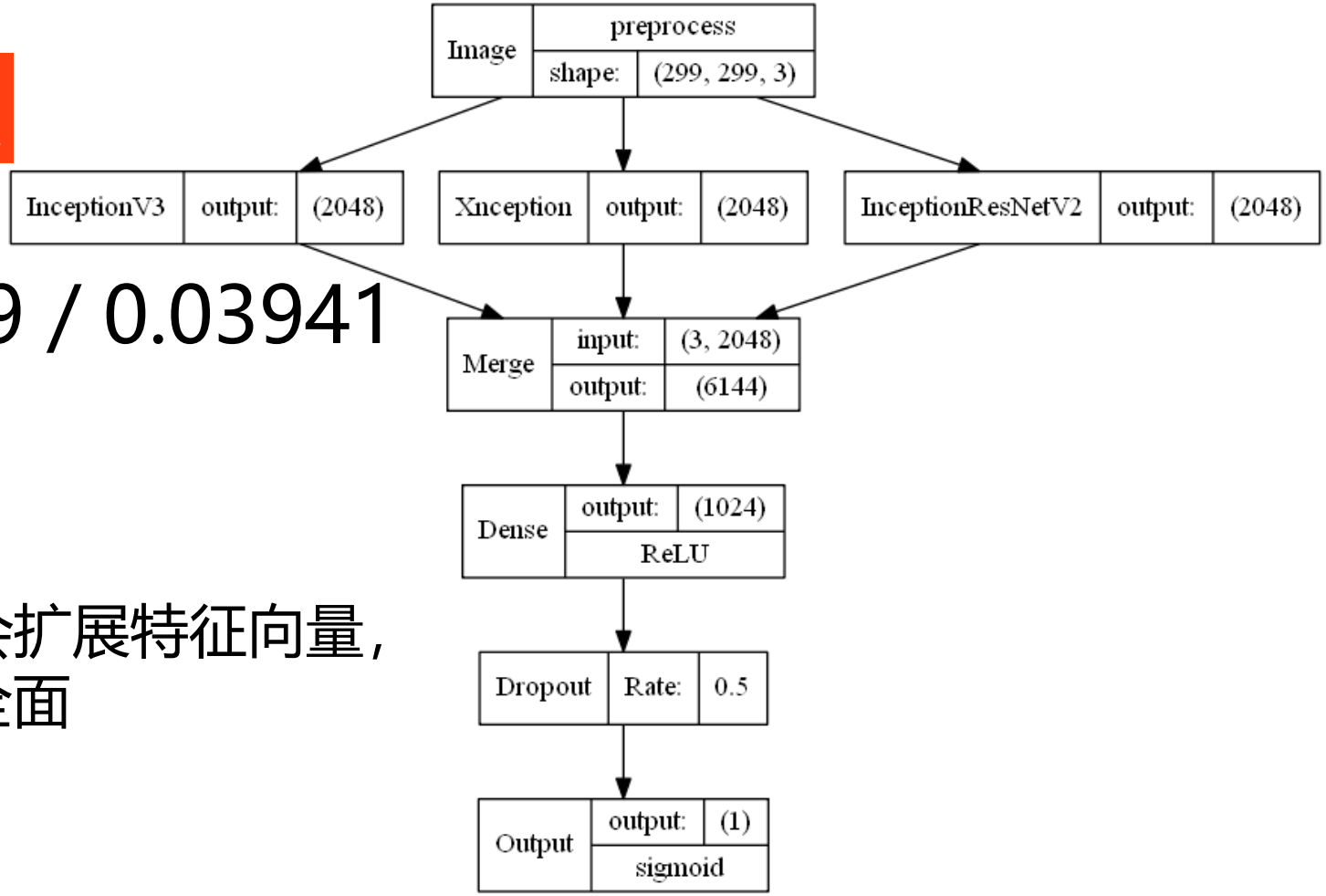


- 分析
  - 使用的SVM分类器输出的是分类结果，而不是概率值，分类错误时受到Log Loss的惩罚更严重
  - FNN表达能力更强

多模型+FNN

- Log Loss: 0.04099 / 0.03941
  - 18th / 15th

- 分析
  - 每增加一个模型，就会扩展特征向量，使提取到的特征更加全面



多模型+FNN+模型融合

- Log Loss: 0.03842
  - 12th, Top1%

- 分析
  - 模型融合的方式，比单纯将CNN网络组合在一起更有效
  - 可以为多模型构建不同的分类器，更加灵活
    - InceptionV3→20epochs
    - Xception→12epochs
    - InceptionResNetV2→12epochs

误差分析

- 预测值在0.2~0.8之间，共134张的测试集图片
  - 错误数据
  - 目标过小
  - 图片质量低
  - 不常见的角度
  - ……


Part 6
改进方法

Part Six 后续改进方法讨论

- 在数据预处理方面，由于训练集图像质量和分辨率不统一，将尝试物体检测的办法，即把猫狗检测出来再切图，用此方法扩大数据集可能会提高准确率

- 网络调优

- 除神经网络之外，在已得到的特征向量基础上尝试传统数据挖掘方法（如随机森林等）