4-4 更多例子

王中雷

厦门大学王亚南经济研究院和经济学院, 2025

内容摘要

1. MobileNet

2. YOLO

内容摘要

1. MobileNet

2. YOLO

MobileNet

1．一个计算效率较高，能够部署在手机等小型设备上的网络结构

- 基于一个“能够在小型计算设备上实现高效计算”的结构
- 该结构为“depth-wise separable convolutions”

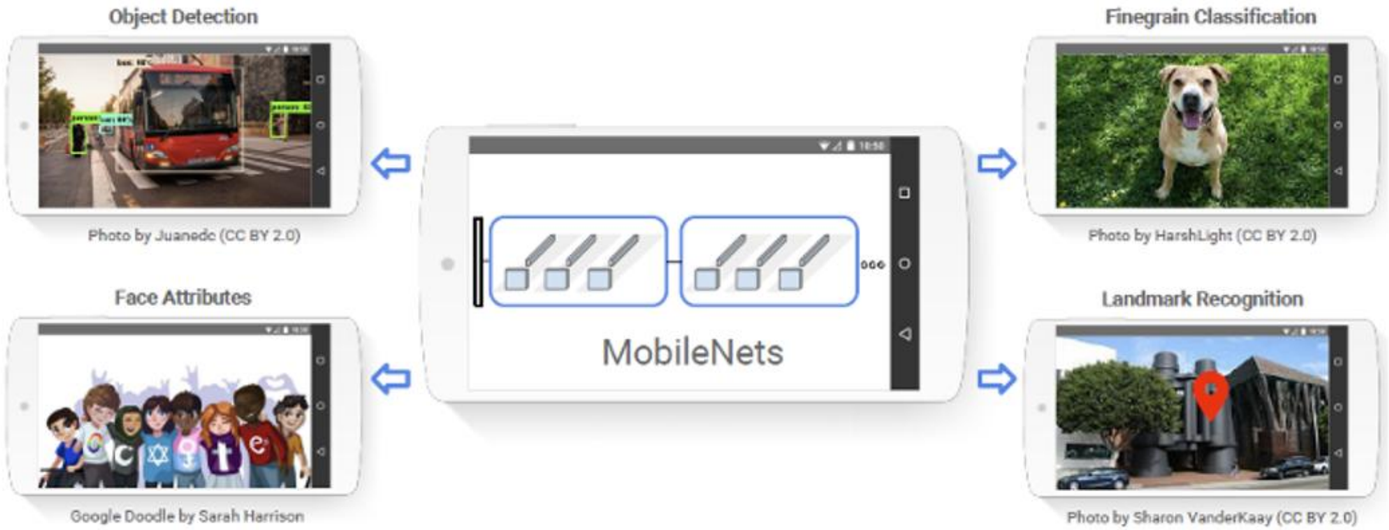
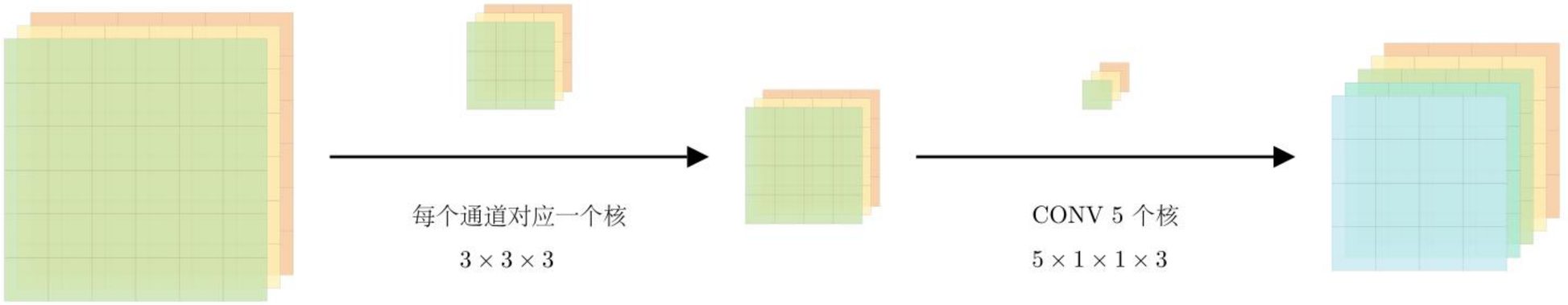


Figure 1. MobileNet models can be applied to various recognition tasks for efficient on device intelligence.

[Howard, A.G., Zhu, M., Chen, B. et al., (2017) MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications]

研究动机



输入

$\mathbb{R}^{6\times6\times3}$

输出

$\mathbb{R}^{4\times4\times5}$

1. 只有$3 \times 3 \times 3 + 3 \times 5$=42 个模型参数

YOLO

1．到目前为止，我们只讨论了分类问题

2．在实际问题中，物体的位置识别也很重要

3．接下来，我们同时解决物体分类以及位置识别两个任务

YOLO

[https://www.superannotate.com/blog/yolo-object-detection]

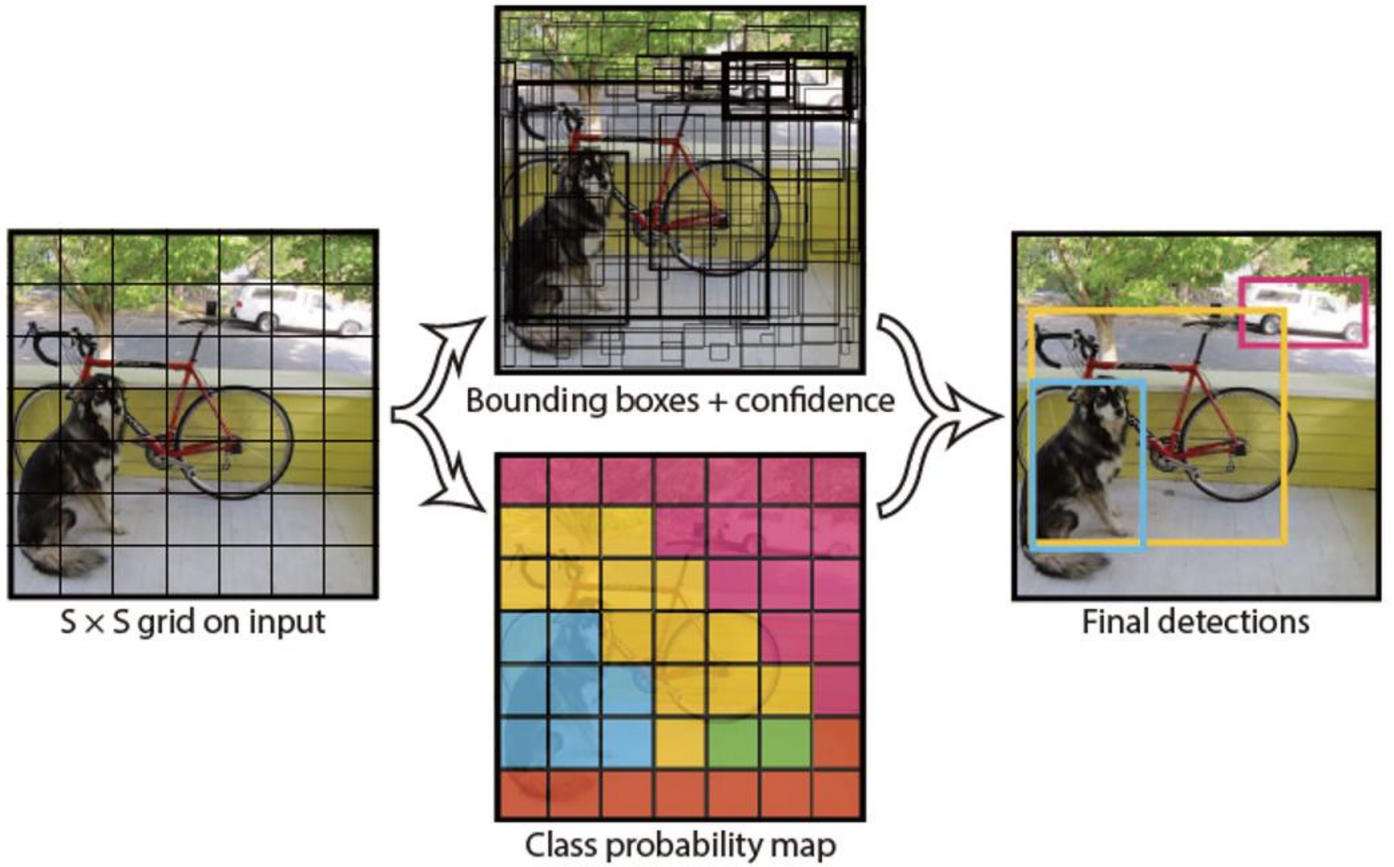
YOLO

1. 我们需要完成C 类的分类问题，并且完成对应物体的位置识别
2. YOLO (You Only Look Once) 为本问题的解决提供了一个很好的思路
3. 我们只介绍 YOLO v1 (Redmon et al., 2016)
4. 更多介绍请参见 https://www.bilibili.com/video/BV1JT411j7MR?p=2

YOLO v1

1. 将图片划分为（不相交的）$S \times S$ 个区域
2. 对于每个区域，考虑B 个边界框（bounding boxes）
3. 对于每个区域，我们基于 IoU 指标选择一个具有“代表性”的边界框
4. 最小化一个特定的代价函数
5. 利用 non-max supression 算法完成物体的位置界定

YOLO v1



[Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016), You Only Look Once: Unified, Real-Time Object Detection, CVPR, 779–788]

YOLO v1

1．将图片划分为（不相交的）$S \times S$ 个区域

YOLO v1

1．对于每个区域，考虑两个（备选）边界框

- 对于每一个物体，我们利用 IoU 指标，只用一个边界框对其进行代表
- 每个边界框由五个元素组成
  - x, y: 边界框相对于对应区域大小的中心坐标
  - w, h: 边界框相对于整个图片大小的宽和高
  - $confidence$: 同时用于衡量边界框包含某物体以及对应分类准确性的信心指标.

YOLO v1

1. IoU 是 “intersection over union” 的缩写
2. 在训练过程中，我们已知物体的位置和标签，我们可以选择一个具有较大 IoU 的边界框
3. 并且在这个基础上，针对相应目标，对边界框进行微调

YOLO v1

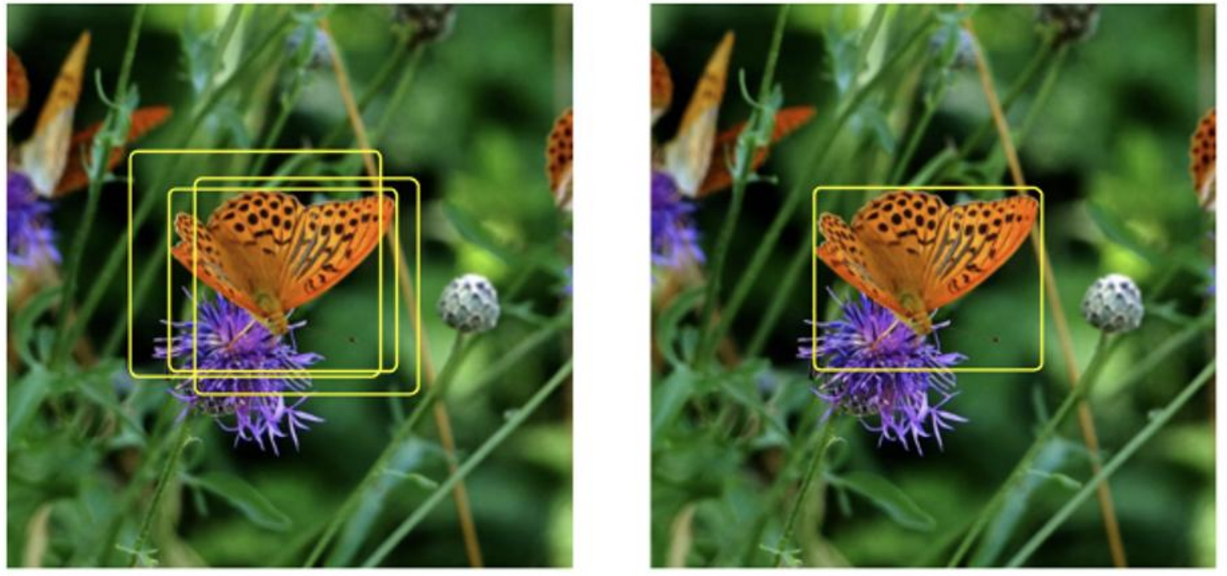
1．基于$\lambda_{\text{coord}}=5$ 以及$\lambda_{\text{noobj}}=0.5$ 的代价函数为

$$
\begin{aligned}
&\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \\
&+ \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right] \\
&+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i \right)^2 \\
&+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i \right)^2 \\
&+ \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} \left(p_i(c) - \hat{p}_i(c) \right)^2 \qquad (3)
\end{aligned}
$$

YOLO v1

1．对于统一物体，首先扔掉所有小于某阈值的边界框

2．在剩余的重合的边界框中，我们只保留 confidence 最大的那个，并且扔掉 IoU 超过某阈值的其他边界框



[https://www.oreilly.com/library/view/practical-machine-learning/9781098102357/ch04.html]