Hypernet 框架

（Towards Accurate Region Proposal Generation and Joint Object Detection.）[1]

1.Abstract

Hypernet 整体框架和 Faster 类似，先通过堆积的卷基层提取特征，再把特征传到 Region Proposal Generation 网络产生 Proposal，接着提供 Proposal 给 Detection Net 做进一步地回归和分类。所不同的是，Hypernet 选取了多个 Feature Map 组合成更具有表现力的特征，叫做 Hyper Feature。

2.Analysis

Hyper Feature

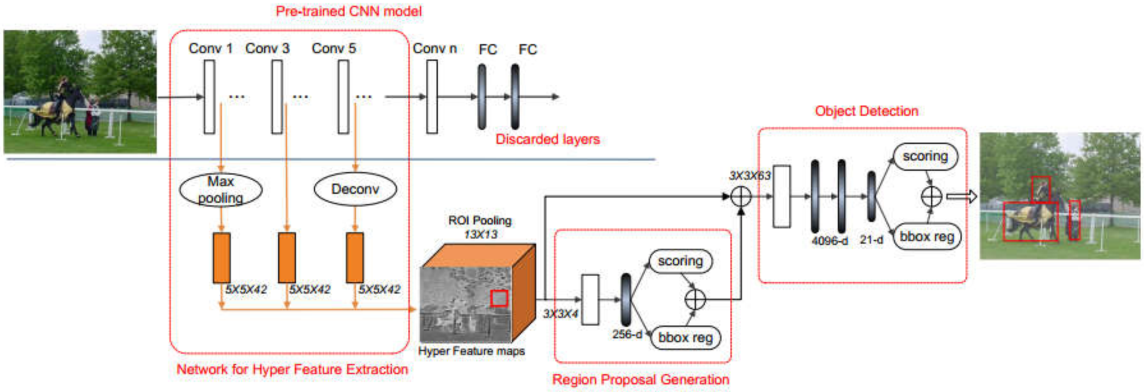

图 12 Framework of Hypernet

如上图所示，框架采取了三个特征图，由于经过多次卷积，各个 Feature Map 的 resolution 不一样，无法直接把它们 concatenate 起来，作者的方法是把 lower layer 做 2*2 的 Max pooling，Middle layer 保持不变，Higher layer 做 Deconvolution，这样最后三类特征图的大小一致，就可以 concatenate 起来。

相比 SSD，这种做法让人感觉是比较好的。SSD 也涉及到多 resolution Feature Map 的处理，它所采取的方法是把特征图 flatten 成向量再 concat，这种方法是很粗糙的，会丧失特征的对应关系，所以 SSD 最终的训练结果也不好。

Joint Training

Algorithm 1 HyperNet training process. After 6 steps, the proposal and detection modules form a unified network.

Step 1: Pre-train a deep CNN model for initializing basic layers in **Step 2** and **Step 3**.
Step 2: Train HyperNet for region proposal generation.
Step 3: Train HyperNet for object detection using region proposals obtained from **Step 2**.
Step 4: Fine-tune HyperNet for region proposal generation sharing Hyper Feature layers trained in **Step 3**.
Step 5: Fine-tune HyperNet for object detection using region proposals obtained from **Step 4**, with shared Hyper Feature layers fixed.
Step 6: Output the unified HyperNet jointly trained in **Step 4** and **Step 5** as the final model.

Method to train Hypernet

训练方法和 Faster 类似。

Speed Up

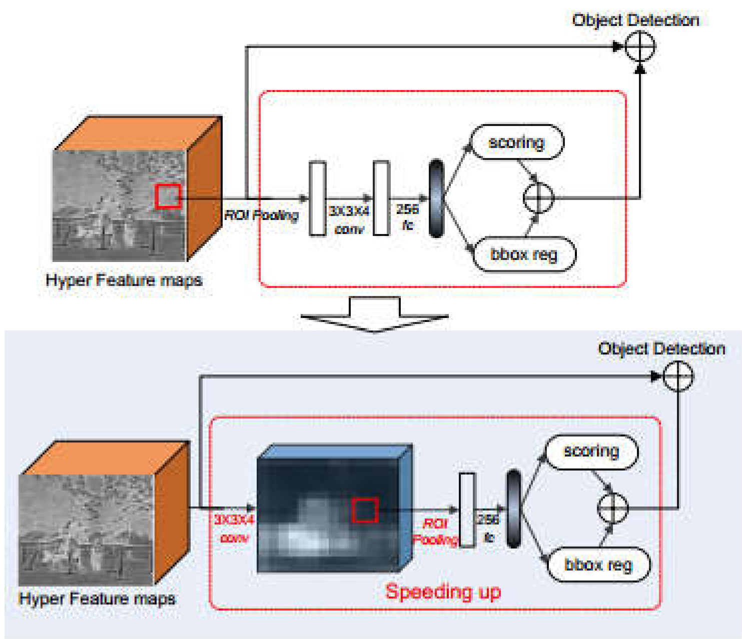

图 13 Speed up the network

在测试的时候，由于需要产生大约 30K 个 roi，每个 roi 都通过后续的 3*3*4 的卷积层很耗费时间，所以可以将该卷积层提前到 ROIpooling 之前，这样后面的 roi 就不需要再做 3*3*4 的卷积，测试 mAP 并没有损失太多。

3.Result

Table1 Result on Pascal_voc 2007

Approach	mAP	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv
Fast R-CNN	70.0	77.0	78.1	69.3	59.4	38.3	81.6	78.6	86.7	42.8	78.8	68.9	84.7	82.0	76.6	69.9	31.8	70.1	74.8	80.4	70.4
Faster R-CNN	73.2	76.5	79.0	70.9	65.5	52.1	**83.1**	84.7	86.4	52.0	**81.9**	65.7	84.8	84.6	77.5	76.7	38.8	73.6	73.9	**83.0**	72.6
HyperNet(AlexNet)	65.9	70.8	75.2	58.2	57.7	40.5	77.6	76.9	74.9	41.3	71.8	66.9	73.7	79.8	75.9	70.9	35.2	62.4	69.2	74.9	63.6
HyperNet	**76.3**	**77.4**	**83.3**	75.0	**69.1**	62.4	**83.1**	**87.4**	**87.4**	**57.1**	79.8	71.4	**85.1**	**85.1**	**80.0**	**79.1**	**51.2**	**79.1**	**75.7**	80.9	**76.5**
HyperNet-SP	74.8	77.3	82.0	**75.4**	64.1	**63.5**	82.5	**87.4**	86.6	55.1	79.3	**71.5**	81.4	84.2	77.6	78.4	45.5	77.4	73.2	78.7	74.8

Table2 Result on Pascal_voc 2012

Approach	mAP	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv
Fast R-CNN	68.4	82.3	78.4	70.8	52.3	38.7	77.8	71.6	**89.3**	44.2	73.0	55.0	**87.5**	80.5	80.8	72.0	35.1	68.3	**65.7**	80.4	64.2
Faster R-CNN	70.4	**84.9**	**79.8**	**74.3**	53.9	49.8	77.5	75.9	88.5	45.6	**77.1**	**55.3**	86.9	**81.7**	80.9	79.6	40.1	72.6	60.9	**81.2**	61.5
NoC	68.8	82.8	79.0	71.6	52.3	**53.7**	74.1	69.0	84.9	46.9	74.3	53.1	85.0	81.3	79.5	72.2	38.9	72.4	59.5	76.7	**68.1**
HyperNet	**71.4**	84.2	78.5	73.6	**55.6**	**53.7**	**78.7**	**79.8**	87.7	**49.6**	74.9	52.1	86.0	**81.7**	**83.3**	**81.8**	**48.6**	**73.5**	59.4	79.9	65.7
HyperNet-SP	71.3	84.1	78.3	73.3	55.5	53.6	78.6	79.6	87.5	49.5	74.9	52.1	85.6	81.6	83.2	81.6	48.4	73.2	59.3	79.7	65.6