Bottom-Up Higher-Resolution Networks for Multi-Person Pose Estimation

用于多人姿态估计的自下而上的高分辨率网络

多人姿势估计的两个方向

- 自上而下（自顶向下）top-down
- 自下而上（自底向上）bottom-up

Top-down

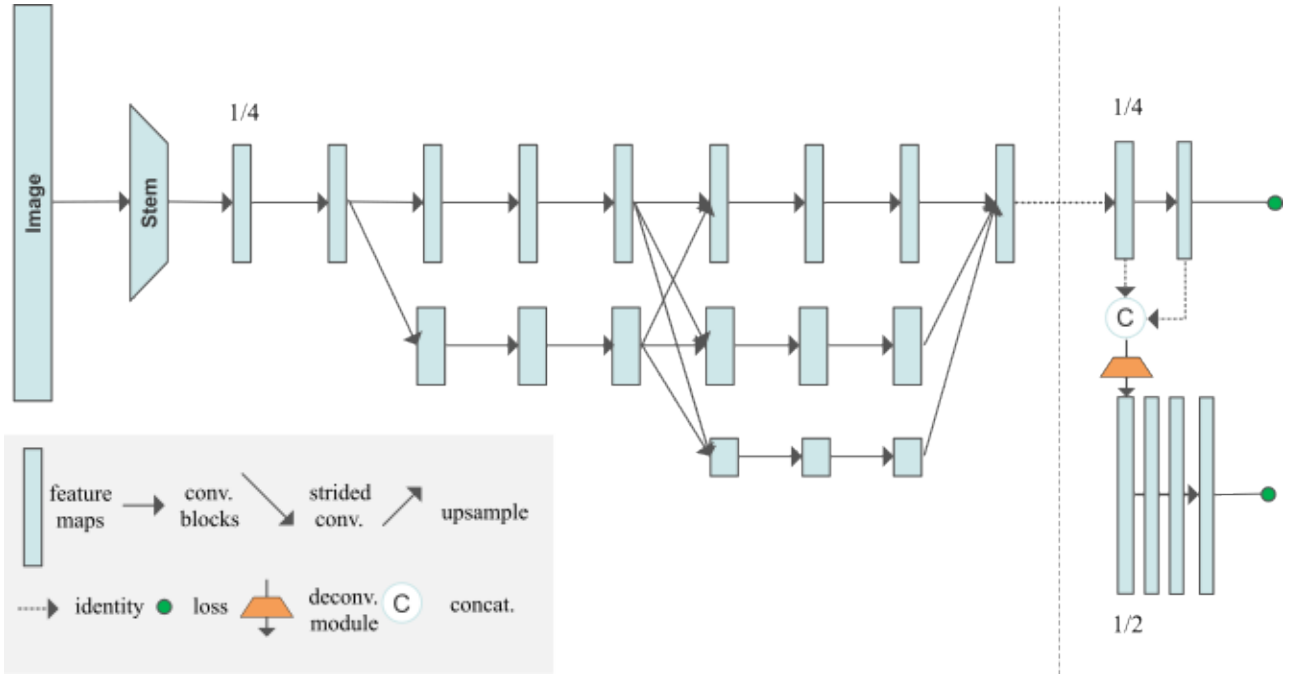
- 自上而下（自顶向下）：先检测单个人，再在单个人里面检测关键点（单人姿势估计）。

bottom-up

- 自下而上（自底向上）：先检测图像中所有的关键点，再将关键点聚类，产生不同个体。

Higher-Resolution Network
高分辨率网络

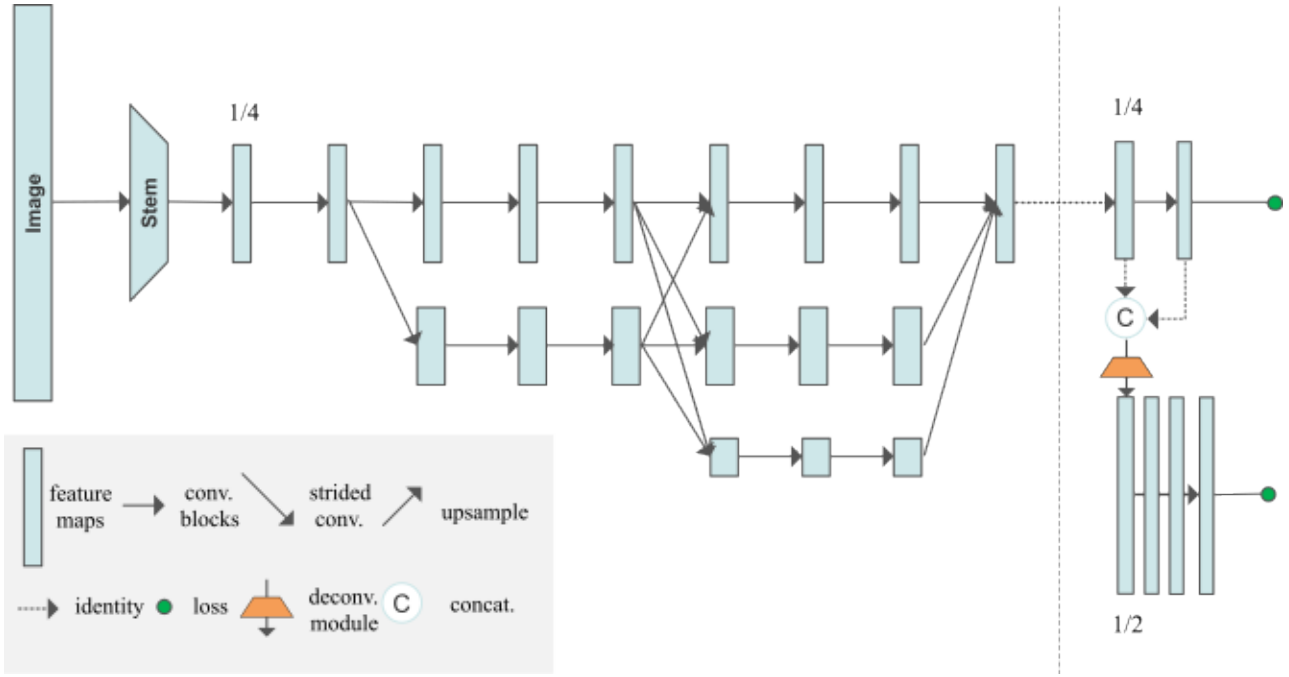
HigherHRNet使用hrnet作为backbone。



Hrnet从第一层开始使用一个高分辨率的分支。

Higher-Resolution Network
高分辨率网络

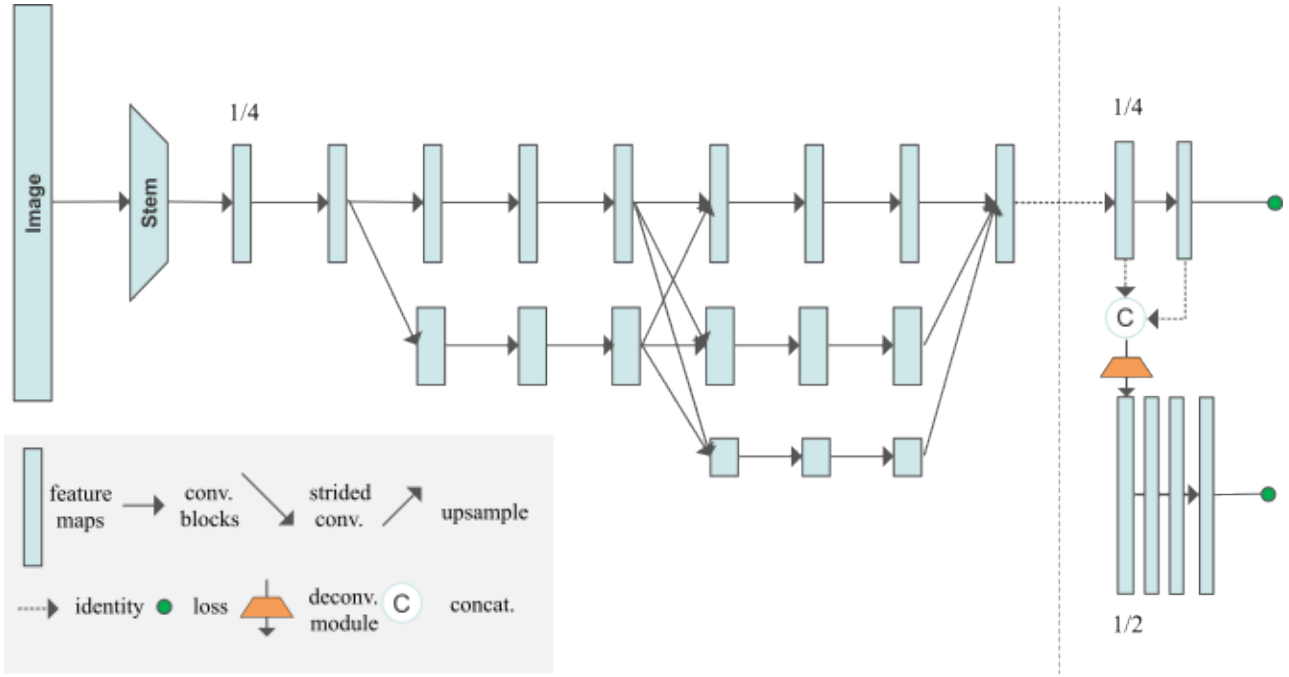
HigherHRNet使用hrnet作为backbone。



在接下来的每一层，一个分辨率为当前分支最低分辨率的二分之一的新分支，并行添加到当前分支。

Higher-Resolution Network
高分辨率网络

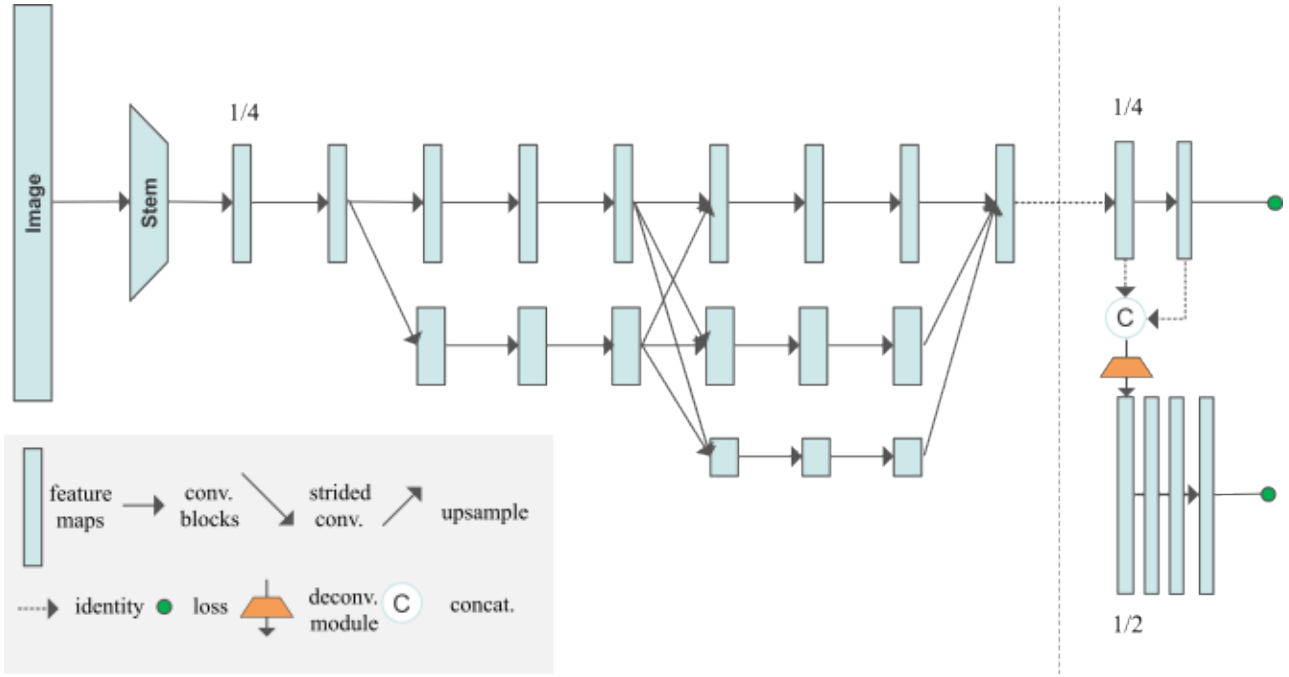
HigherHRNet使用hrnet作为backbone。



上图是三个并行分支的网络结构。

Higher-Resolution Network
高分辨率网络

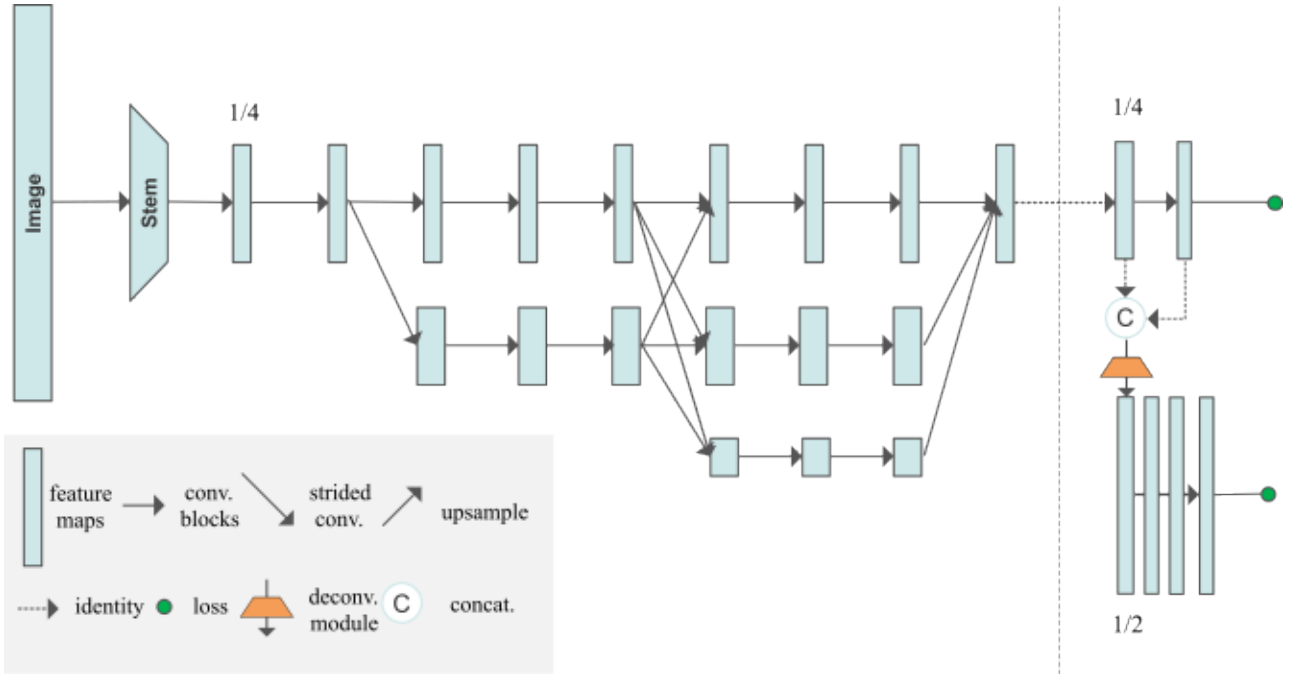
HigherHRNet使用hrnet作为backbone。



在1/4输入图片分辨率的特征图上使用反卷积提高分辨率，提升中等尺寸的人体姿势估计。（coco数据集关键点检测任务中没有小尺寸人物的标注）

Higher-Resolution Network
高分辨率网络

HigherHRNet使用hrnet作为backbone。



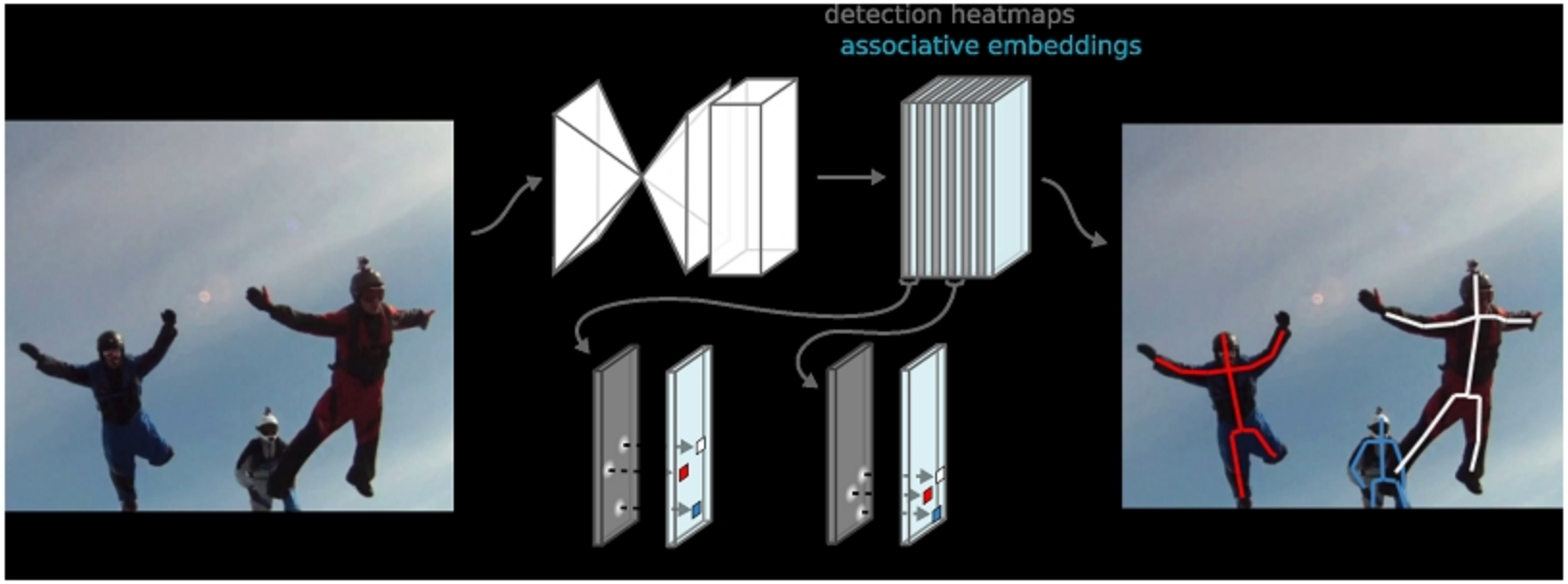
反卷积层的添加，使hrnet变为HigherHRNet。

聚类
grouping

- 通过关联嵌入（associative embedding）一种类似热力图的方式，预测关键点的标签（tag），通过计算关键点的标签的l2距离，对关键点聚类分组。

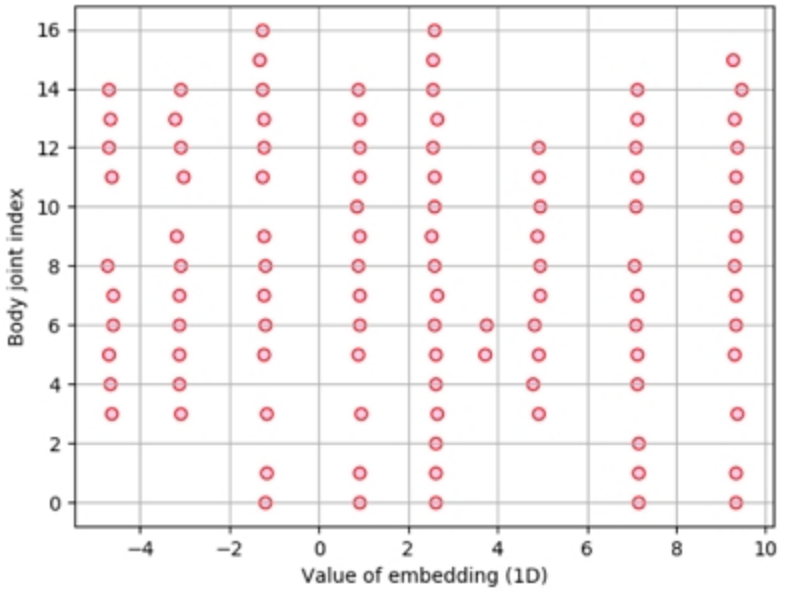
聚类
grouping

- 通过关联嵌入（associative embedding）一种类似热力图的方式，预测关键点的标签（tag），通过计算关键点的标签的l2距离，对关键点聚类分组。

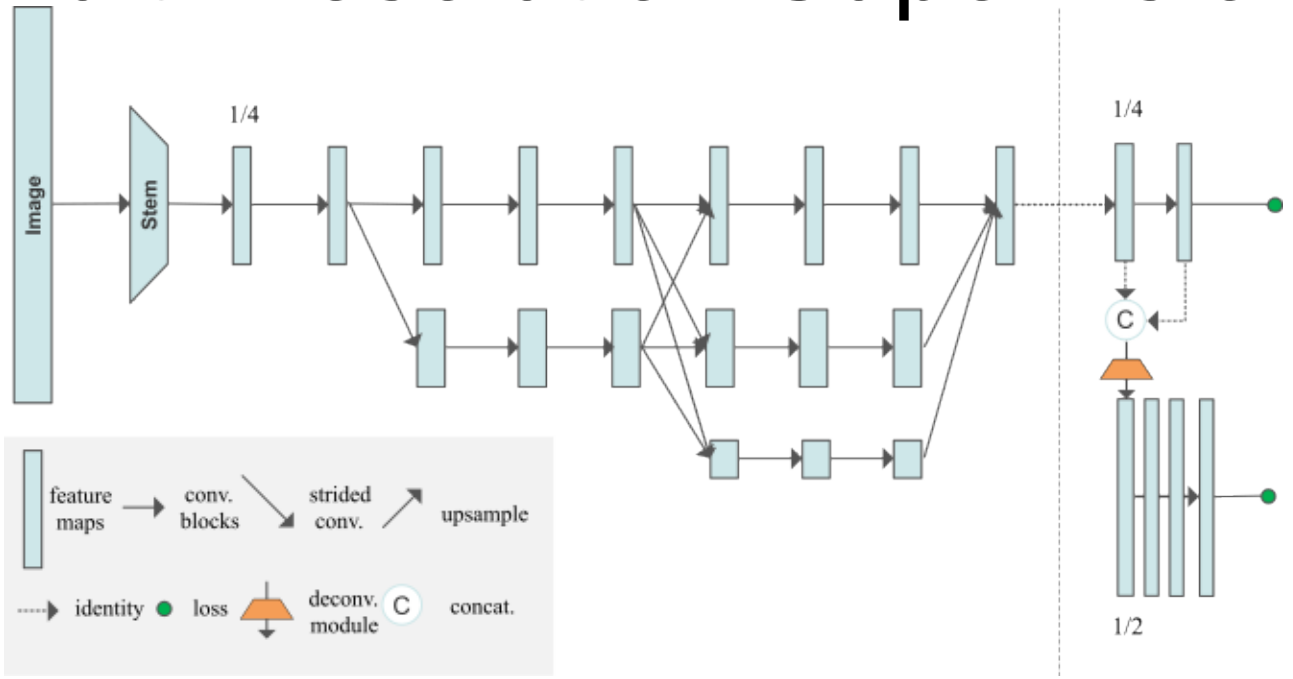


聚类
grouping

- 通过关联嵌入（associative embedding）一种类似热力图的方式，预测关键点的标签（tag），通过计算关键点的标签的l2距离，对关键点聚类分组。

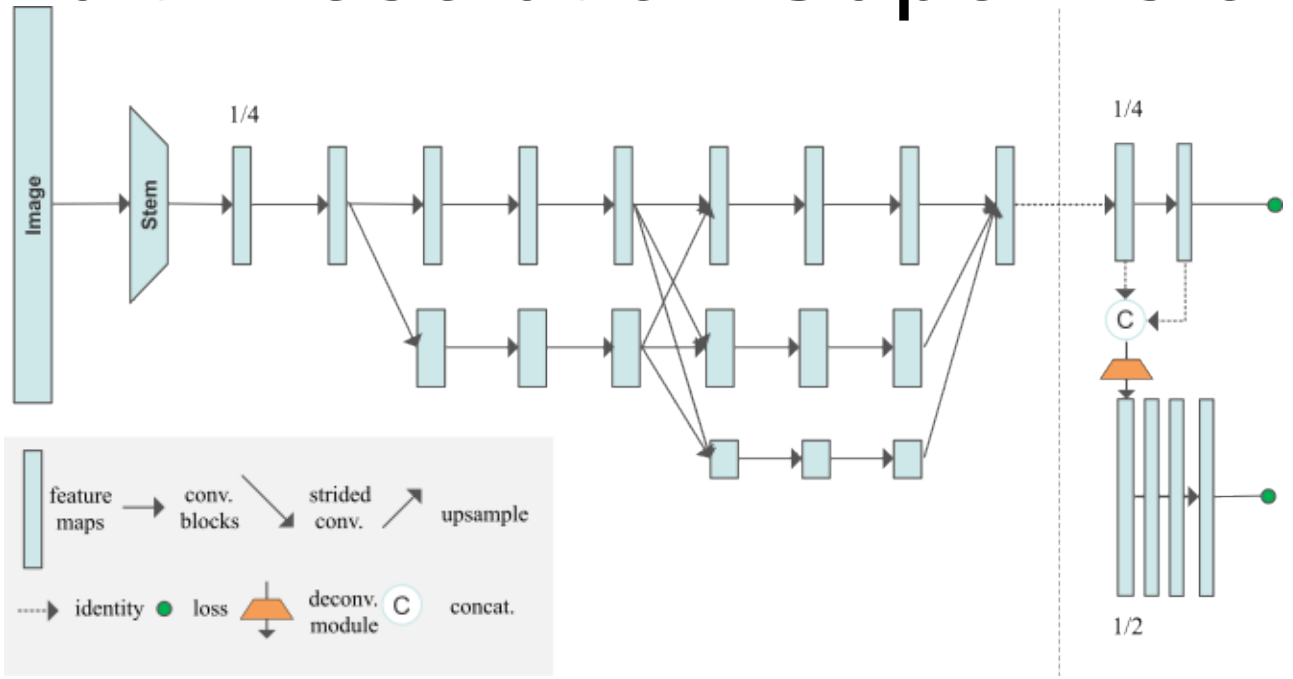


多分辨率监督
Multi-Resolution Supervision



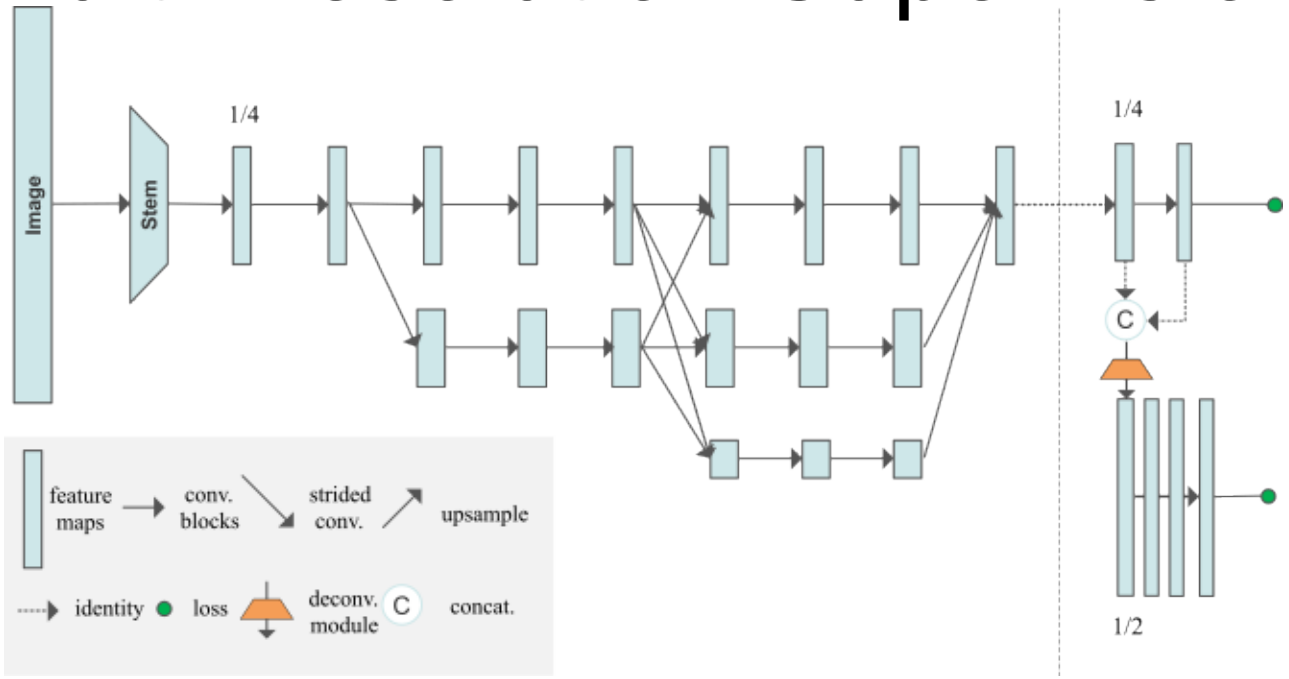
与其它自下而上方法只应用最大分辨率的热力图监督不同，higherhrnet在训练时使用多个分辨率的热力图，去解决尺寸变化。

多分辨率监督
Multi-Resolution Supervision



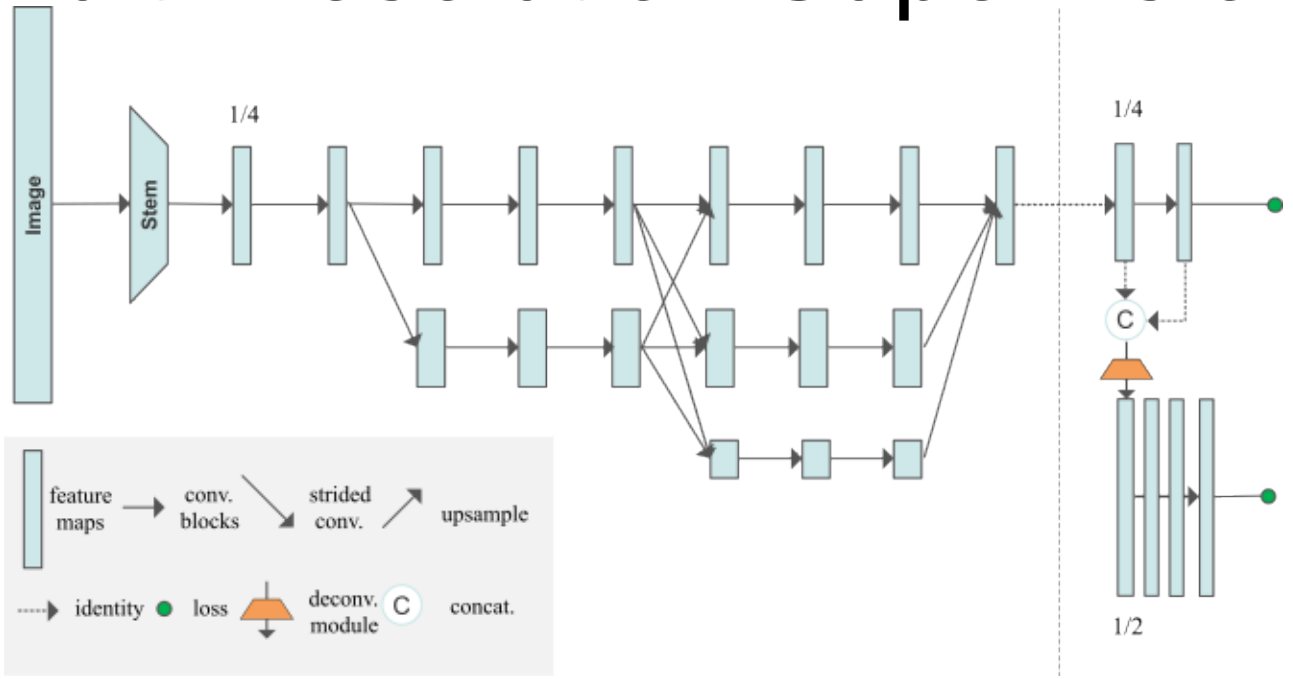
将真实的关键点位置转移到不同分辨率的热力图对应的位置上。但对所有的这些不同分辨率的热力图使用同一标准差高斯核。

多分辨率监督
Multi-Resolution Supervision



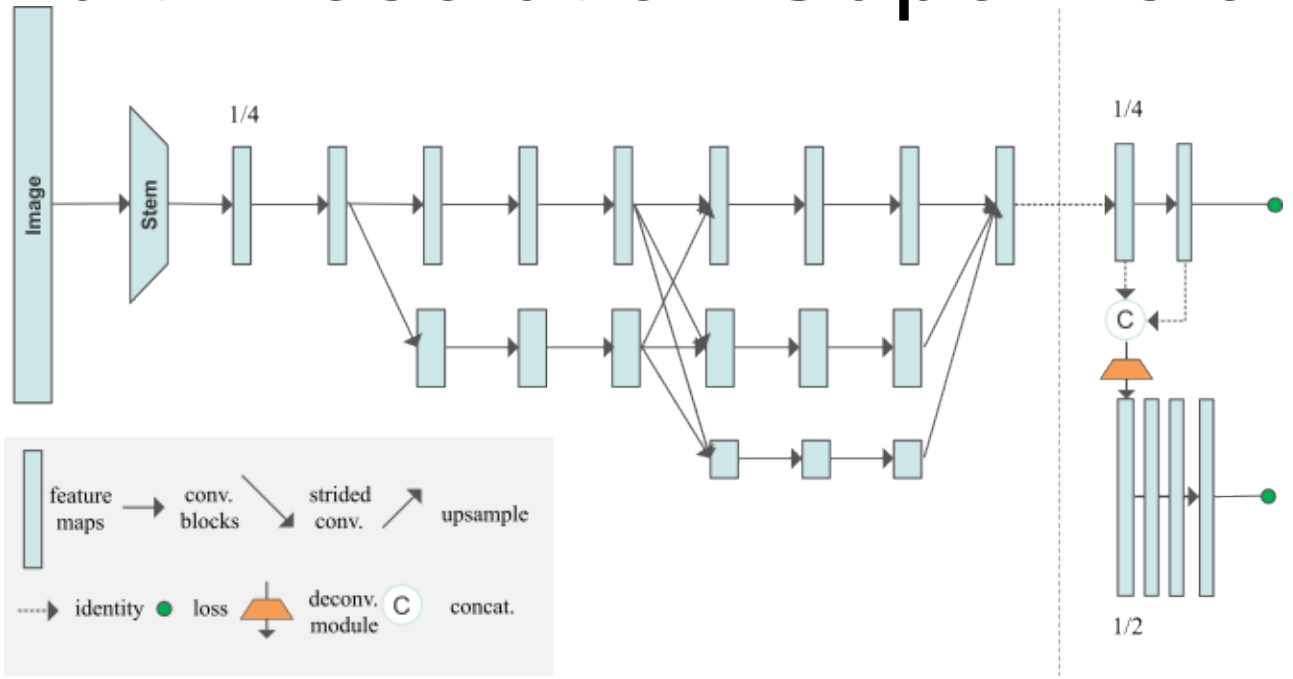
不同分辨率的特征金字塔适用于不同尺度的关键点预测。在更高分辨率的特征图上，为了更精确地定位中等尺寸人物的关键点，需要一个相对较小的标准差。

多分辨率监督
Multi-Resolution Supervision



每个分辨率尺度上，计算推理得到的热力图和对应的真实热力图的均方差。最后将其相加作为热力图loss。

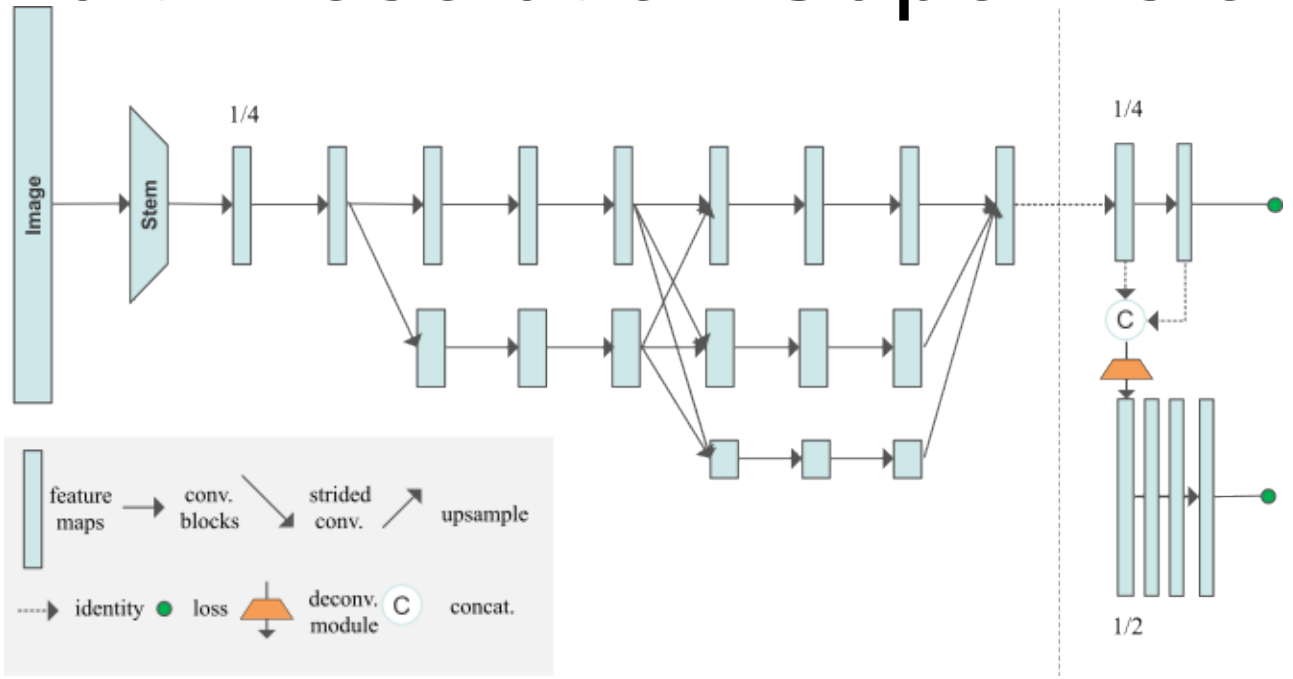
多分辨率监督
Multi-Resolution Supervision



说明：没有将不同尺度的人物目标分配给不同尺度的特征金字塔。

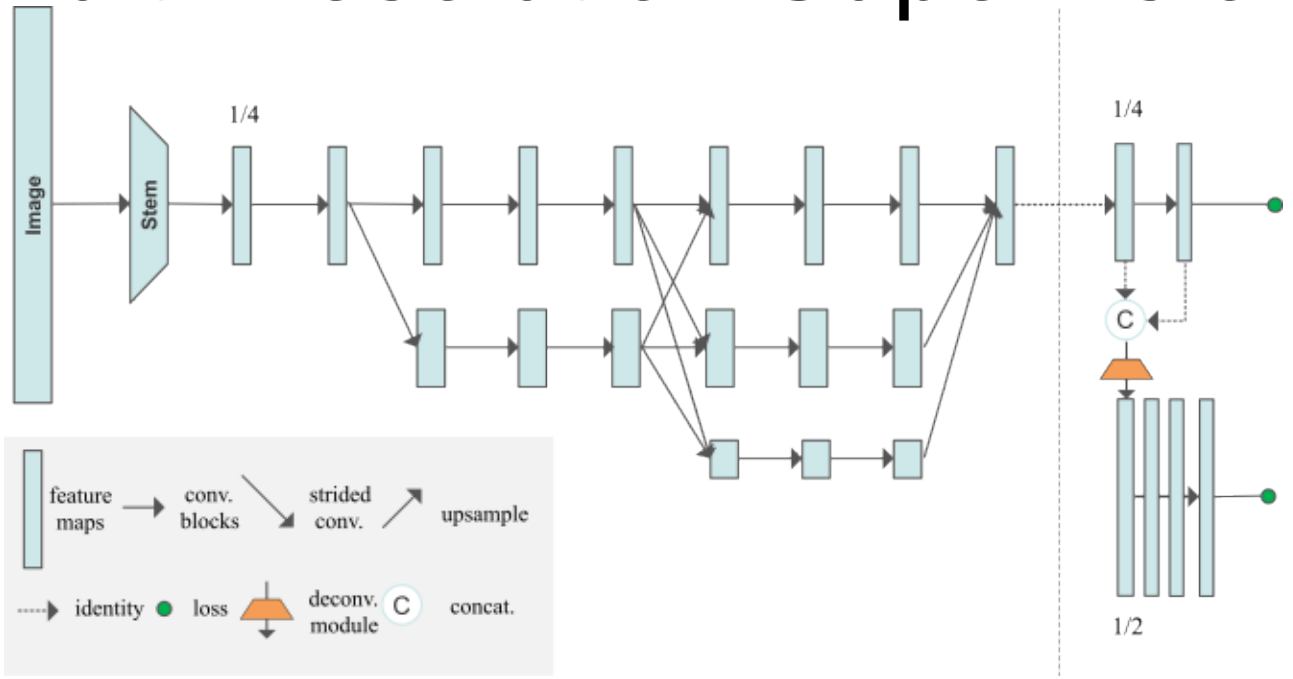
有两点原因。

多分辨率监督
Multi-Resolution Supervision



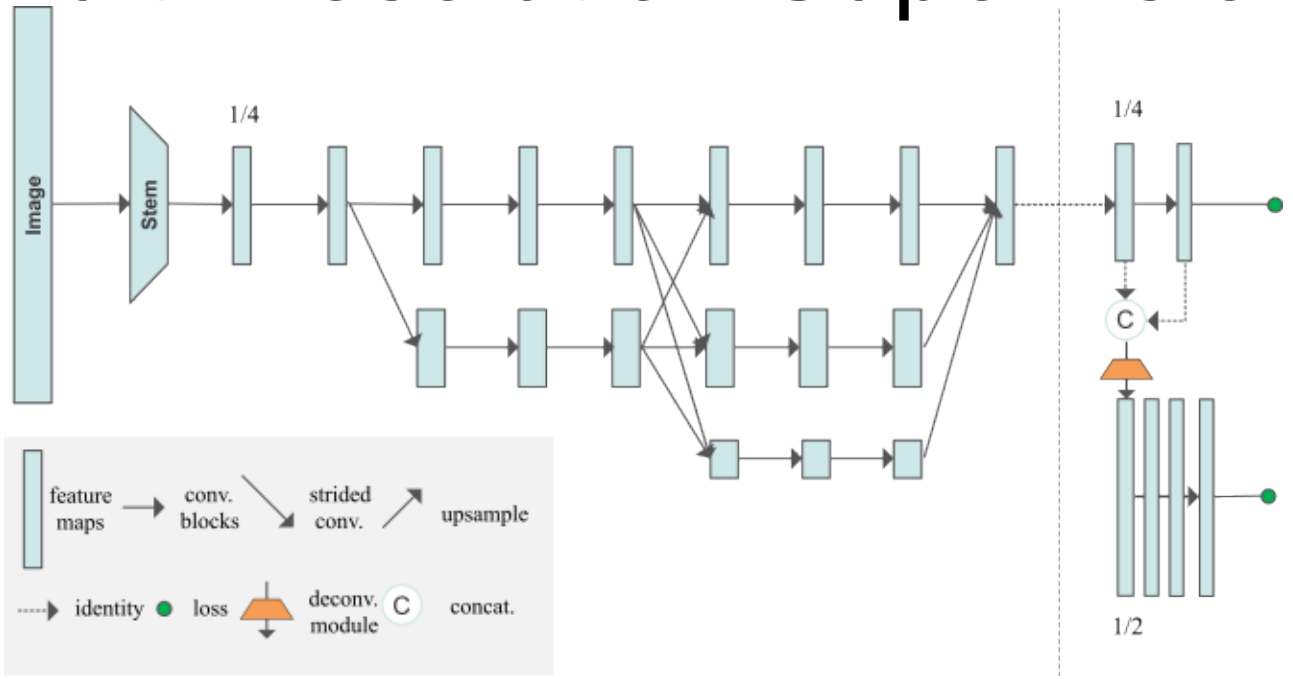
第一点：分配训练目标到不同层的特征金字塔是受到数据集和网络结构的启发。

多分辨率监督
Multi-Resolution Supervision



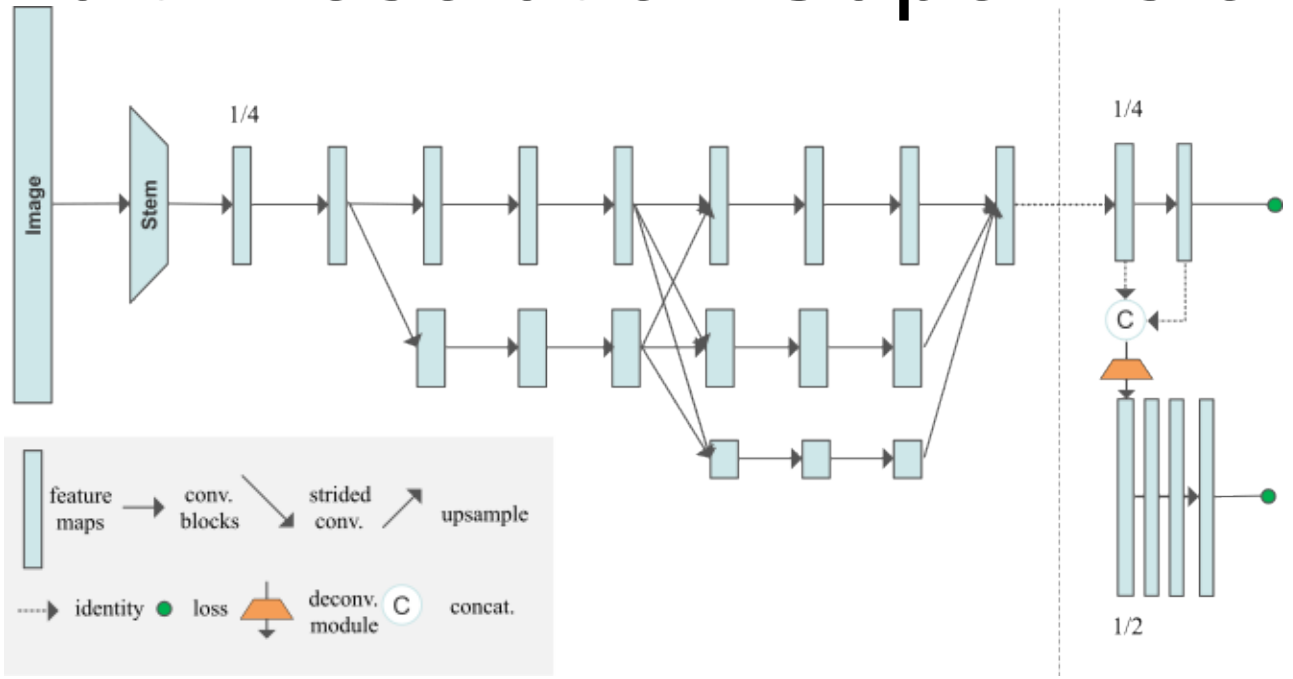
数据集：人的尺寸分布vs所有物体的尺寸分布。
网络结构：higherhrnet 2层 vs fpn 4层。
数据集和网络结构都变化的情况，很难将启发从fpn转移到higherhrnet上。

多分辨率监督
Multi-Resolution Supervision



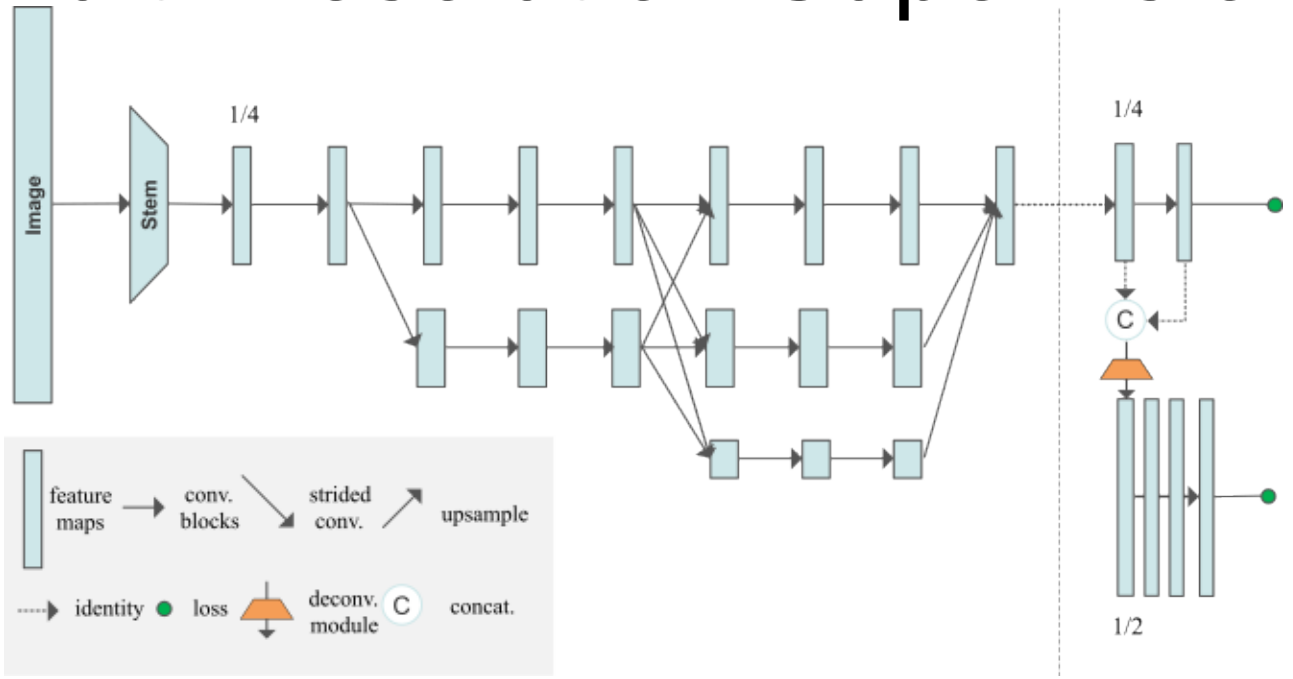
第二点：由于使用了高斯核，真实的关键点目标相互影响。很难通过简单的设置忽略区域来解耦关键点。

多分辨率监督
Multi-Resolution Supervision



与关键点位置的热力图使用所有分辨率不同，标签（tag）的热力图只在低分辨率上预测。

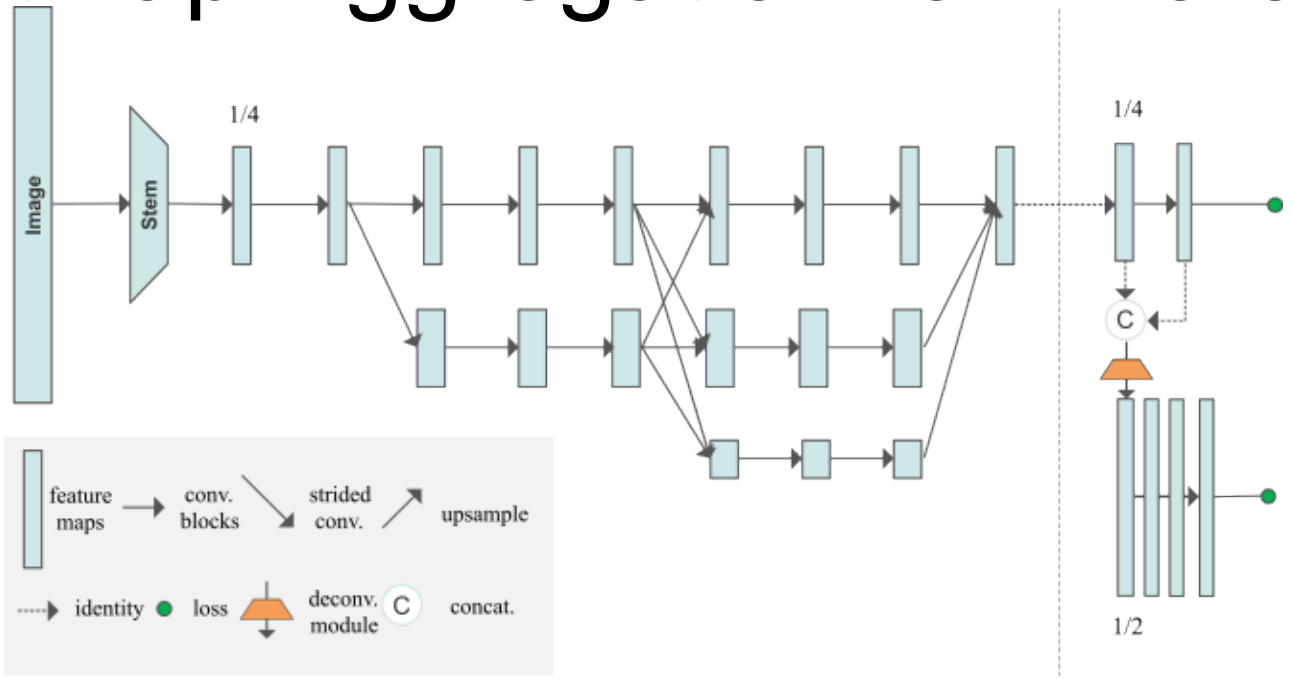
多分辨率监督
Multi-Resolution Supervision



学习标签（tag）的热力图需要全局推理，其更适合在低分辨率上去预测。

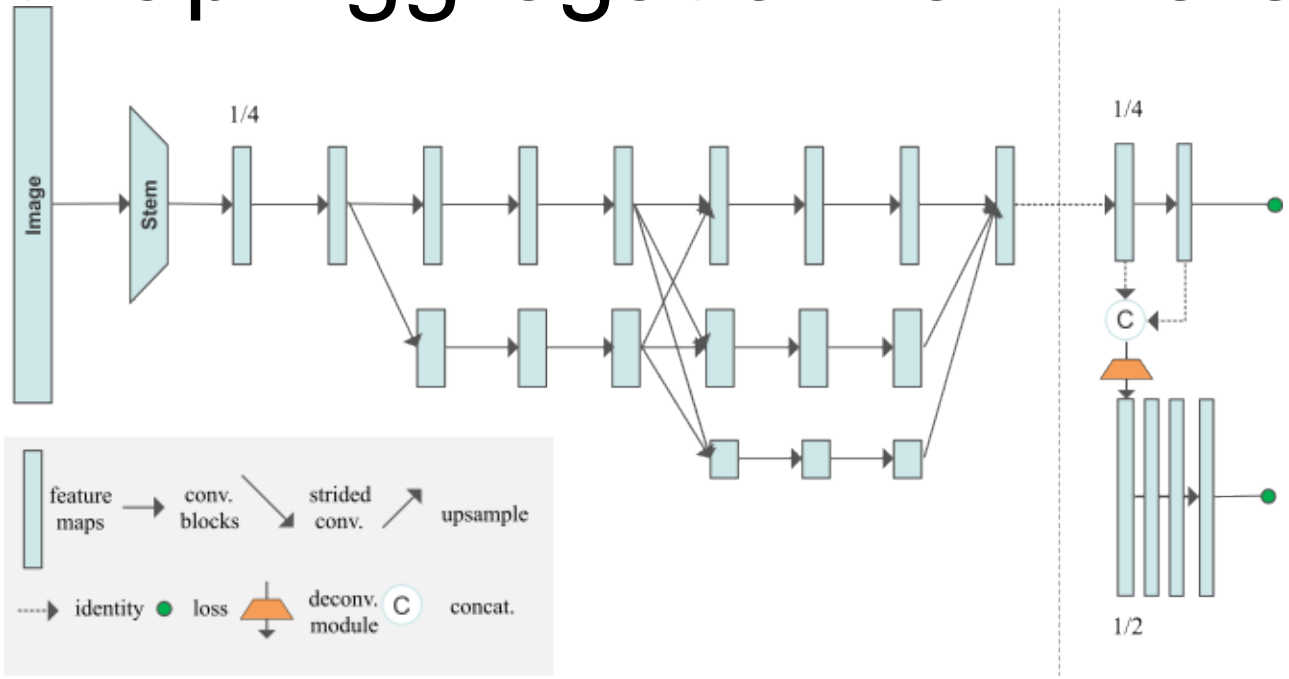
论文作者通过实验发现，更高的分辨率不能把标签（tag）的热力图学得好，甚至会不收敛。

推理时热力图融合
Heatmap Aggregation for Inference



使用上采样双线性差值将不同分辨率热力图恢复到输入图片的分辨率。将这些热力图取平均作为最后的结果。

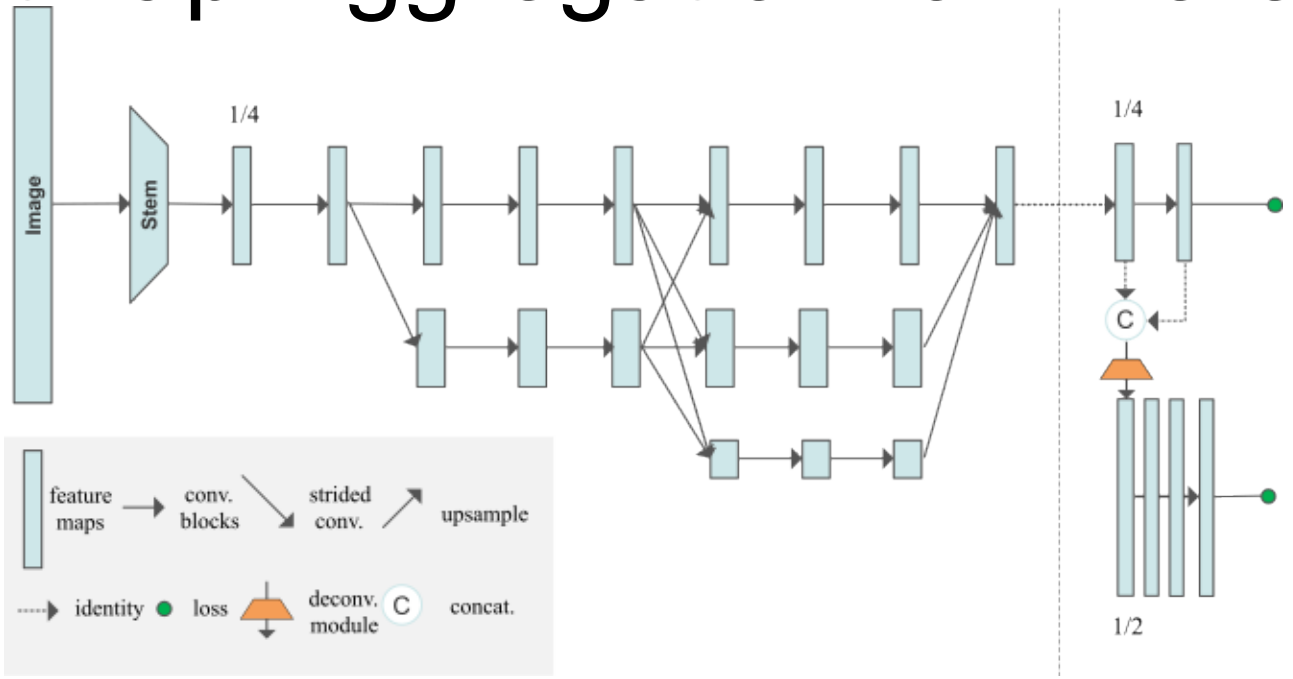
推理时热力图融合
Heatmap Aggregation for Inference



这一做法与以前的方法差别很大。

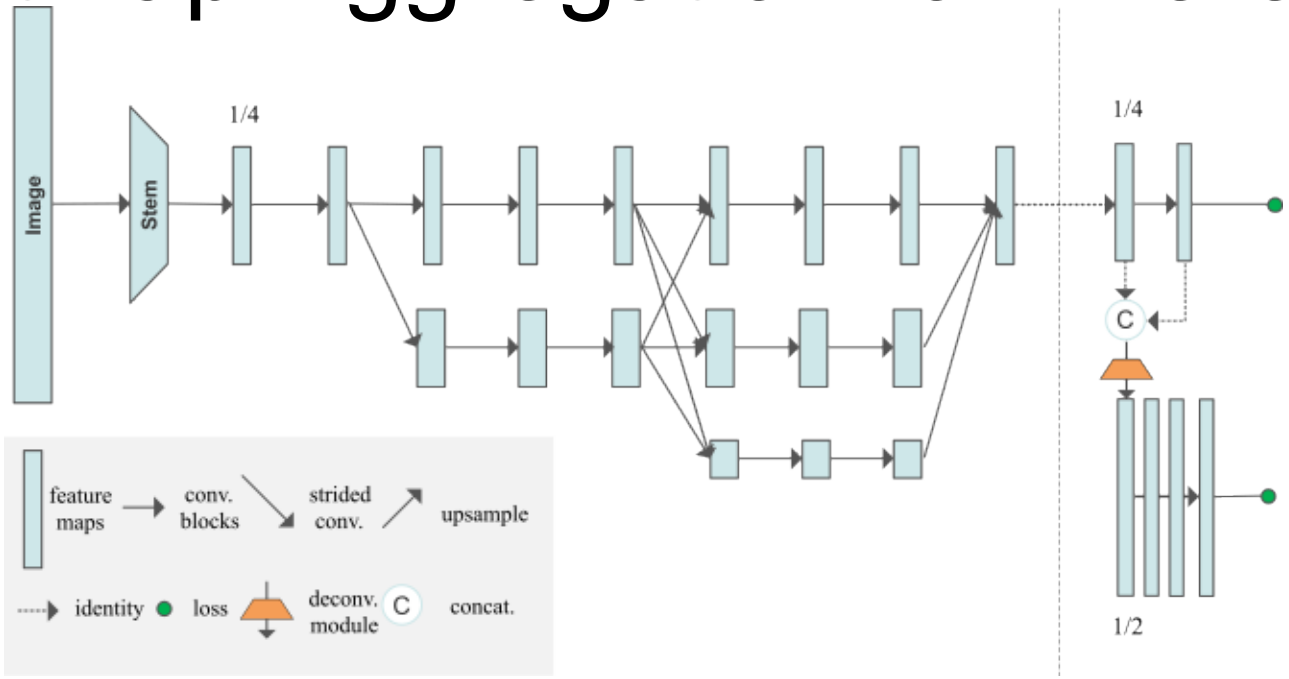
以前的方法只从预测的一个尺寸或一层得到热力图。

推理时热力图融合
Heatmap Aggregation for Inference



这一做法，使用热图聚合,原因是为了实现能尺度感知的姿态估计。Coco关键点数据集中人的尺寸变化大，范围从32x32到128x128。

推理时热力图融合
Heatmap Aggregation for Inference



论文作者发现在higherhrnet中来自不同尺度的热图能更好地捕捉在不同尺度的关键点。

具体的，在低分辨率热力图中小尺寸人物消失的关键点能在高分辨率热力图中恢复。这样实现尺度感知的姿态估计。

Coco Experiments

Method	Backbone	Input size	#Params	GFLOPs	AP	AP^{50}	AP^{75}	AP^{M}	AP^{L}
w/o multi-scale test									
OpenPose [3]†	-	-	-	-	61.8	84.9	67.5	57.1	68.2
Hourglass [25]	Hourglass	512	277.8M	206.9	56.6	81.8	61.8	49.8	67.0
PersonLab [28]	ResNet-152	1401	68.7M	405.5	66.5	88.0	72.6	62.4	72.3
PifPaf [18]	-	-	-	-	66.7	-	-	62.4	72.9
Bottom-up HRNet‡	HRNet-w32	512	28.5M	38.9	64.1	86.3	70.4	57.4	73.9
Ours	HRNet-w32	512	28.6M	47.9	66.4	87.5	72.8	61.2	74.2
Ours	HRNet-w48	640	63.8M	154.3	**68.4**	**88.2**	**75.1**	**64.4**	**74.2**
w/ multi-scale test									
Hourglass [25]	Hourglass	512	277.8M	206.9	63.0	85.7	68.9	58.0	70.4
Hourglass [25]†	Hourglass	512	277.8M	206.9	65.5	86.8	72.3	60.6	72.6
PersonLab [28]	ResNet-152	1401	68.7M	405.5	68.7	89.0	75.4	64.1	75.5
Ours	HRNet-w48	640	63.8M	154.3	**70.5**	**89.3**	**77.2**	**66.6**	**75.8**

† Indicates using refinement.

‡ Our implementation, not reported in [33]

Table 1. Comparisons with bottom-up methods on the COCO2017 test-dev set. All GFLOPs are calculated at single-scale. For PersonLab [28], we only calculate its backbone's #Params and GFLOPs. Top: w/o multi-scale test. Bottom: w/ multi-scale test. *It is worth noting that our results are achieved without refinement.*

Results on COCO2017 test-dev dataset

Coco Experiments

Method	Backbone	Input size	#Params	GFLOPs	AP	AP^{50}	AP^{75}	AP^{M}	AP^{L}
w/o multi-scale test									
OpenPose [3]†	-	-	-	-	61.8	84.9	67.5	57.1	68.2
Hourglass [25]	Hourglass	512	277.8M	206.9	56.6	81.8	61.8	49.8	67.0
PersonLab [28]	ResNet-152	1401	68.7M	405.5	66.5	88.0	72.6	62.4	72.3
PifPaf [18]	-	-	-	-	66.7	-	-	62.4	72.9
Bottom-up HRNet‡	HRNet-w32	512	28.5M	38.9	64.1	86.3	70.4	57.4	73.9
Ours	HRNet-w32	512	28.6M	47.9	66.4	87.5	72.8	61.2	74.2
Ours	HRNet-w48	640	63.8M	154.3	**68.4**	**88.2**	**75.1**	**64.4**	**74.2**
w/ multi-scale test									
Hourglass [25]	Hourglass	512	277.8M	206.9	63.0	85.7	68.9	58.0	70.4
Hourglass [25]†	Hourglass	512	277.8M	206.9	65.5	86.8	72.3	60.6	72.6
PersonLab [28]	ResNet-152	1401	68.7M	405.5	68.7	89.0	75.4	64.1	75.5
Ours	HRNet-w48	640	63.8M	154.3	**70.5**	**89.3**	**77.2**	**66.6**	**75.8**

† Indicates using refinement.

‡ Our implementation, not reported in [33]

Table 1. Comparisons with bottom-up methods on the COCO2017 test-dev set. All GFLOPs are calculated at single-scale. For PersonLab [28], we only calculate its backbone's #Params and GFLOPs. Top: w/o multi-scale test. Bottom: w/ multi-scale test. *It is worth noting that our results are achieved without refinement.*

Results on COCO2017 test-dev dataset

Hrnet本身在自下而上方法中就作为一个简单而强大的基准线（64.1 ap）。

Coco Experiments

Method	Backbone	Input size	#Params	GFLOPs	AP	AP^{50}	AP^{75}	AP^{M}	AP^{L}
w/o multi-scale test									
OpenPose [3]†	-	-	-	-	61.8	84.9	67.5	57.1	68.2
Hourglass [25]	Hourglass	512	277.8M	206.9	56.6	81.8	61.8	49.8	67.0
PersonLab [28]	ResNet-152	1401	68.7M	405.5	66.5	88.0	72.6	62.4	72.3
PifPaf [18]	-	-	-	-	66.7	-	-	62.4	72.9
Bottom-up HRNet‡	HRNet-w32	512	28.5M	38.9	64.1	86.3	70.4	57.4	73.9
Ours	HRNet-w32	512	28.6M	47.9	66.4	87.5	72.8	61.2	74.2
Ours	HRNet-w48	640	63.8M	154.3	**68.4**	**88.2**	**75.1**	**64.4**	**74.2**
w/ multi-scale test									
Hourglass [25]	Hourglass	512	277.8M	206.9	63.0	85.7	68.9	58.0	70.4
Hourglass [25]†	Hourglass	512	277.8M	206.9	65.5	86.8	72.3	60.6	72.6
PersonLab [28]	ResNet-152	1401	68.7M	405.5	68.7	89.0	75.4	64.1	75.5
Ours	HRNet-w48	640	63.8M	154.3	**70.5**	**89.3**	**77.2**	**66.6**	**75.8**

† Indicates using refinement.

‡ Our implementation, not reported in [33]

Table 1. Comparisons with bottom-up methods on the COCO2017 test-dev set. All GFLOPs are calculated at single-scale. For PersonLab [28], we only calculate its backbone's #Params and GFLOPs. Top: w/o multi-scale test. Bottom: w/ multi-scale test. *It is worth noting that our results are achieved without refinement.*

Results on COCO2017 test-dev dataset

Hrnet单尺寸测试的表现超过了hourglass的多尺寸测试的结果。

Coco Experiments

Method	Backbone	Input size	#Params	GFLOPs	AP	AP^{50}	AP^{75}	AP^{M}	AP^{L}
w/o multi-scale test									
OpenPose [3]†	-	-	-	-	61.8	84.9	67.5	57.1	68.2
Hourglass [25]	Hourglass	512	277.8M	206.9	56.6	81.8	61.8	49.8	67.0
PersonLab [28]	ResNet-152	1401	68.7M	405.5	66.5	88.0	72.6	62.4	72.3
PifPaf [18]	-	-	-	-	66.7	-	-	62.4	72.9
Bottom-up HRNet‡	HRNet-w32	512	28.5M	38.9	64.1	86.3	70.4	57.4	73.9
Ours	HRNet-w32	512	28.6M	47.9	66.4	87.5	72.8	61.2	74.2
Ours	HRNet-w48	640	63.8M	154.3	**68.4**	**88.2**	**75.1**	**64.4**	**74.2**
w/ multi-scale test									
Hourglass [25]	Hourglass	512	277.8M	206.9	63.0	85.7	68.9	58.0	70.4
Hourglass [25]†	Hourglass	512	277.8M	206.9	65.5	86.8	72.3	60.6	72.6
PersonLab [28]	ResNet-152	1401	68.7M	405.5	68.7	89.0	75.4	64.1	75.5
Ours	HRNet-w48	640	63.8M	154.3	**70.5**	**89.3**	**77.2**	**66.6**	**75.8**

† Indicates using refinement.

‡ Our implementation, not reported in [33]

Table 1. Comparisons with bottom-up methods on the COCO2017 test-dev set. All GFLOPs are calculated at single-scale. For PersonLab [28], we only calculate its backbone's #Params and GFLOPs. Top: w/o multi-scale test. Bottom: w/ multi-scale test. *It is worth noting that our results are achieved without refinement.*

Results on COCO2017 test-dev dataset

Higherhrnet（66.4 ap）超过hrnet （64.1 ap）的表现2.3 ap。

Coco Experiments

Method	Backbone	Input size	#Params	GFLOPs	AP	AP^{50}	AP^{75}	AP^{M}	AP^{L}
w/o multi-scale test									
OpenPose [3]†	-	-	-	-	61.8	84.9	67.5	57.1	68.2
Hourglass [25]	Hourglass	512	277.8M	206.9	56.6	81.8	61.8	49.8	67.0
PersonLab [28]	ResNet-152	1401	68.7M	405.5	66.5	88.0	72.6	62.4	72.3
PifPaf [18]	-	-	-	-	66.7	-	-	62.4	72.9
Bottom-up HRNet‡	HRNet-w32	512	28.5M	38.9	64.1	86.3	70.4	57.4	73.9
Ours	HRNet-w32	512	28.6M	47.9	66.4	87.5	72.8	61.2	74.2
Ours	HRNet-w48	640	63.8M	154.3	**68.4**	**88.2**	**75.1**	**64.4**	**74.2**
w/ multi-scale test									
Hourglass [25]	Hourglass	512	277.8M	206.9	63.0	85.7	68.9	58.0	70.4
Hourglass [25]†	Hourglass	512	277.8M	206.9	65.5	86.8	72.3	60.6	72.6
PersonLab [28]	ResNet-152	1401	68.7M	405.5	68.7	89.0	75.4	64.1	75.5
Ours	HRNet-w48	640	63.8M	154.3	**70.5**	**89.3**	**77.2**	**66.6**	**75.8**

† Indicates using refinement.

‡ Our implementation, not reported in [33]

Table 1. Comparisons with bottom-up methods on the COCO2017 test-dev set. All GFLOPs are calculated at single-scale. For PersonLab [28], we only calculate its backbone's #Params and GFLOPs. Top: w/o multi-scale test. Bottom: w/ multi-scale test. *It is worth noting that our results are achieved without refinement.*

Results on COCO2017 test-dev dataset

Higherhrnet在多尺度测试中达到了70.5 ap，大幅度超过现有的自下而上的方法。

Coco Experiments

Method	AP	AP^{50}	AP^{75}	AP^{M}	AP^{L}	AR
Top-down methods						
Mask-RCNN [12]	63.1	87.3	68.7	57.8	71.4	-
G-RMI [29]	64.9	85.5	71.3	62.3	70.0	69.7
Integral Pose Regression [34]	67.8	88.2	74.8	63.9	74.0	-
G-RMI + extra data [29]	68.5	87.1	75.5	65.8	73.3	73.3
CPN [9]	72.1	91.4	80.0	68.7	77.2	78.5
RMPE [11]	72.3	89.2	79.1	68.0	78.6	-
CFN [14]	72.6	86.1	69.7	78.3	64.1	-
CPN (ensemble) [9]	73.0	91.7	80.9	69.5	78.1	79.0
SimpleBaseline [36]	73.7	91.9	81.1	70.3	80.0	79.0
HRNet-W48 [33]	75.5	92.5	83.3	71.9	81.5	80.5
HRNet-W48 + extra data [33]	77.0	92.7	84.5	73.4	83.1	82.0
Bottom-up methods						
OpenPose* [3]	61.8	84.9	67.5	57.1	68.2	66.5
Hourglass*+ [25]	65.5	86.8	72.3	60.6	72.6	70.2
PifPaf [18]	66.7	-	-	62.4	72.9	-
SPM [27]	66.9	88.5	72.9	62.6	73.1	-
PersonLab+ [28]	68.7	89.0	75.4	64.1	75.5	75.4
Ours: HigherHRNet-W48+	70.5	89.3	77.2	66.6	75.8	74.9

Table 2. Comparisons with both top-down and bottom-up methods on COCO2017 test-dev dataset. * means using refinement. + means using multi-scale test.

- both bottom-up and top-down methods on the COCO2017 test -dev dataset

Coco Experiments

Method	AP	AP^{50}	AP^{75}	AP^{M}	AP^{L}	AR
Top-down methods						
Mask-RCNN [12]	63.1	87.3	68.7	57.8	71.4	-
G-RMI [29]	64.9	85.5	71.3	62.3	70.0	69.7
Integral Pose Regression [34]	67.8	88.2	74.8	63.9	74.0	-
G-RMI + extra data [29]	68.5	87.1	75.5	65.8	73.3	73.3
CPN [9]	72.1	91.4	80.0	68.7	77.2	78.5
RMPE [11]	72.3	89.2	79.1	68.0	78.6	-
CFN [14]	72.6	86.1	69.7	78.3	64.1	-
CPN (ensemble) [9]	73.0	91.7	80.9	69.5	78.1	79.0
SimpleBaseline [36]	73.7	91.9	81.1	70.3	80.0	79.0
HRNet-W48 [33]	75.5	92.5	83.3	71.9	81.5	80.5
HRNet-W48 + extra data [33]	77.0	92.7	84.5	73.4	83.1	82.0
Bottom-up methods						
OpenPose* [3]	61.8	84.9	67.5	57.1	68.2	66.5
Hourglass*+ [25]	65.5	86.8	72.3	60.6	72.6	70.2
PifPaf [18]	66.7	-	-	62.4	72.9	-
SPM [27]	66.9	88.5	72.9	62.6	73.1	-
PersonLab+ [28]	68.7	89.0	75.4	64.1	75.5	75.4
Ours: HigherHRNet-W48+	70.5	89.3	77.2	66.6	75.8	74.9

Table 2. Comparisons with both top-down and bottom-up methods on COCO2017 test-dev dataset. * means using refinement. + means using multi-scale test.

- both bottom-up and top-down methods on the COCO2017 test-dev dataset
- Higherhrnet进一步缩小了自下而上和自上而下方法之间的性能差距

消融实验hrnet vs higherhrnet Ablation Experiments

Method	Feat. stride/resolution	AP	AP^M	AP^L
HRNet	4/128	64.4	57.1	75.6
HigherHRNet	2/256	66.9	61.0	75.7
HigherHRNet	1/512	66.5	61.1	74.9

Table 3. Ablation study of HRNet *vs*. HigherRNet on COCO2017 val dataset. Using one deconvolution module for HigherHRNet performs best on the COCO dataset.

- 一个使用特征步长为4的hrnet的自下而上方法ap=64.4。

消融实验hrnet vs higherhrnet Ablation Experiments

Method	Feat. stride/resolution	AP	AP^M	AP^L
HRNet	4/128	64.4	57.1	75.6
HigherHRNet	2/256	66.9	61.0	75.7
HigherHRNet	1/512	66.5	61.1	74.9

Table 3. Ablation study of HRNet *vs*. HigherRNet on COCO2017 val dataset. Using one deconvolution module for HigherHRNet performs best on the COCO dataset.

- 一个使用特征步长为4的hrnet的自下而上方法ap=64.4。
- 添加一个反卷积模块，特征步长为2的higherhrnet获得了2.5 ap的巨大提升（66.9 ap）。

消融实验hrnet vs higherhrnet Ablation Experiments

Method	Feat. stride/resolution	AP	AP^M	AP^L
HRNet	4/128	64.4	57.1	75.6
HigherHRNet	2/256	66.9	61.0	75.7
HigherHRNet	1/512	66.5	61.1	74.9

Table 3. Ablation study of HRNet *vs*. HigherRNet on COCO2017 val dataset. Using one deconvolution module for HigherHRNet performs best on the COCO dataset.

- 一个使用特征步长为4的hrnet的自下而上方法ap=64.4。
- 添加一个反卷积模块，使特征步长为2的higherhrnet获得了2.5 ap的巨大提升（66.9 ap）。
- 提升主要来自于小尺寸的人物目标，apM从57.1提升到61.0。

消融实验hrnet vs higherhrnet Ablation Experiments

Method	Feat. stride/resolution	AP	AP^M	AP^L
HRNet	4/128	64.4	57.1	75.6
HigherHRNet	2/256	66.9	61.0	75.7
HigherHRNet	1/512	66.5	61.1	74.9

Table 3. Ablation study of HRNet *vs*. HigherRNet on COCO2017 val dataset. Using one deconvolution module for HigherHRNet performs best on the COCO dataset.

- 结果展示了，得益于higherhrnet的高分辨率特征图，在小尺寸上表现更好。
- higherhrnet推理时也使用了低分辨率的特征图，大尺寸的人物目标也没有下降。

消融实验hrnet vs higherhrnet Ablation Experiments

Method	Feat. stride/resolution	AP	AP^M	AP^L
HRNet	4/128	64.4	57.1	75.6
HigherHRNet	2/256	66.9	61.0	75.7
HigherHRNet	1/512	66.5	61.1	74.9

Table 3. Ablation study of HRNet *vs*. HigherRNet on COCO2017 val dataset. Using one deconvolution module for HigherHRNet performs best on the COCO dataset.

- 结果说明了预测使用更高的分辨率有利于自下而上的姿势估计。
- 能感知尺度变化的预测是重要的。

消融实验hrnet vs higherhrnet Ablation Experiments

Method	Feat. stride/resolution	AP	AP^M	AP^L
HRNet	4/128	64.4	57.1	75.6
HigherHRNet	2/256	66.9	61.0	75.7
HigherHRNet	1/512	66.5	61.1	74.9

Table 3. Ablation study of HRNet *vs*. HigherRNet on COCO2017 val dataset. Using one deconvolution module for HigherHRNet performs best on the COCO dataset.

- 再增加一个反卷积模块提升特征图分辨率，ap从66.9下降到了66.5。
- 在apM上提升了0.1，在apL上下降了0.8。

消融实验hrnet vs higherhrnet Ablation Experiments

Method	Feat. stride/resolution	AP	AP^M	AP^L
HRNet	4/128	64.4	57.1	75.6
HigherHRNet	2/256	66.9	61.0	75.7
HigherHRNet	1/512	66.5	61.1	74.9

Table 3. Ablation study of HRNet *vs*. HigherRNet on COCO2017 val dataset. Using one deconvolution module for HigherHRNet performs best on the COCO dataset.

论文作者推测是由于特征图尺度和目标尺度失调。

- 更大分辨率的特征图对更小的人物目标的关键点检测有利，但coco数据集中没有考虑小目标的姿势估计，没有小目标样本。
- 因此对coco数据集使用一个反卷积模块，但论文作者指出级联反卷积模块的数量取决于数据集。

消融实验 higherhrnet详解
Ablation Experiments

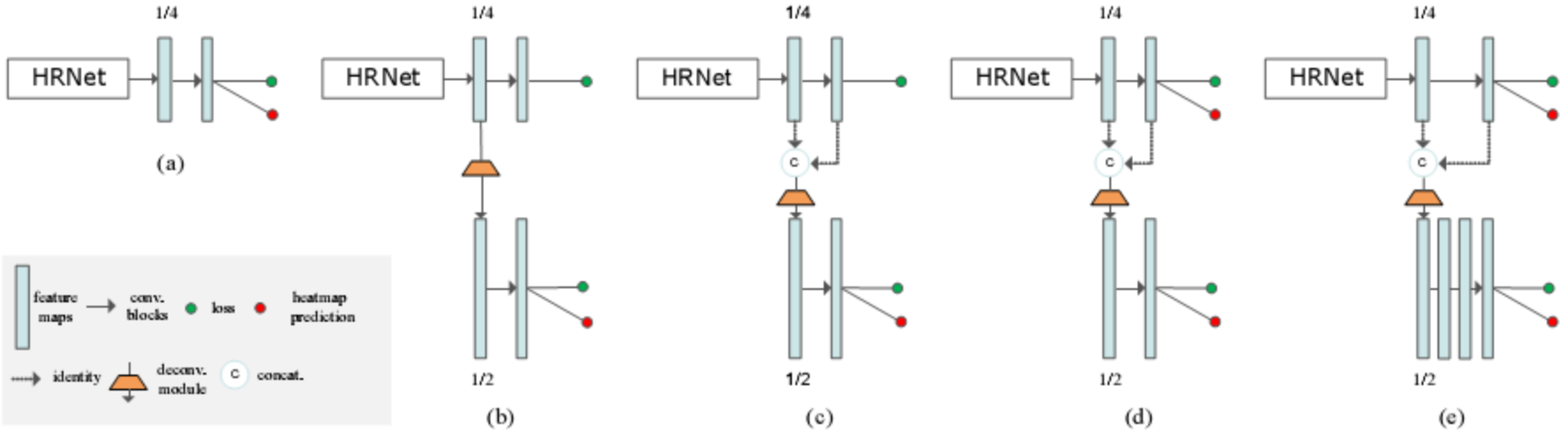


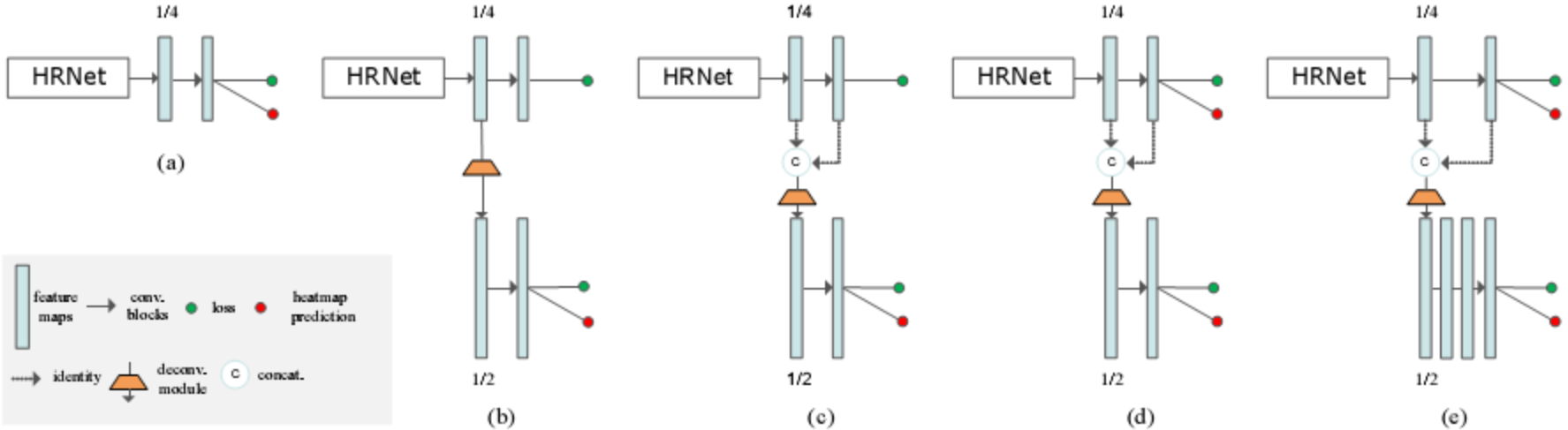
Figure 3. (a) Baseline method using HRNet [33] as backbone. (b) HigherHRNet with multi-resolution supervision (MRS). (c) HigherHRNet with MRS and feature concatenation. (d) HigherHRNet with MRS and feature concatenation. (e) HigherHRNet with MRS, feature concatennation and extra residual blocks. For (d) and (e), heatmap aggregation is used.

	Network	w/ MRS	feature concat.	w/ heatmap aggregation	extra res. blocks	AP	AP^M	AP^L
(a)	HRNet					64.4	57.1	75.6
(b)	HigherHRNet	✓				66.0	60.7	74.2
(c)	HigherHRNet	✓	✓			66.3	60.8	74.0
(d)	HigherHRNet	✓	✓	✓		66.9	61.0	75.7
(e)	HigherHRNet	✓	✓	✓	✓	67.1	61.5	76.1

Table 4. Ablation study of HigherHRNet's components. MSR: multi-resolution supervision. feature concat.: feature concatenation. res. blocks: residual blocks.

图3是作者实验的所有结构。结果在表4中。

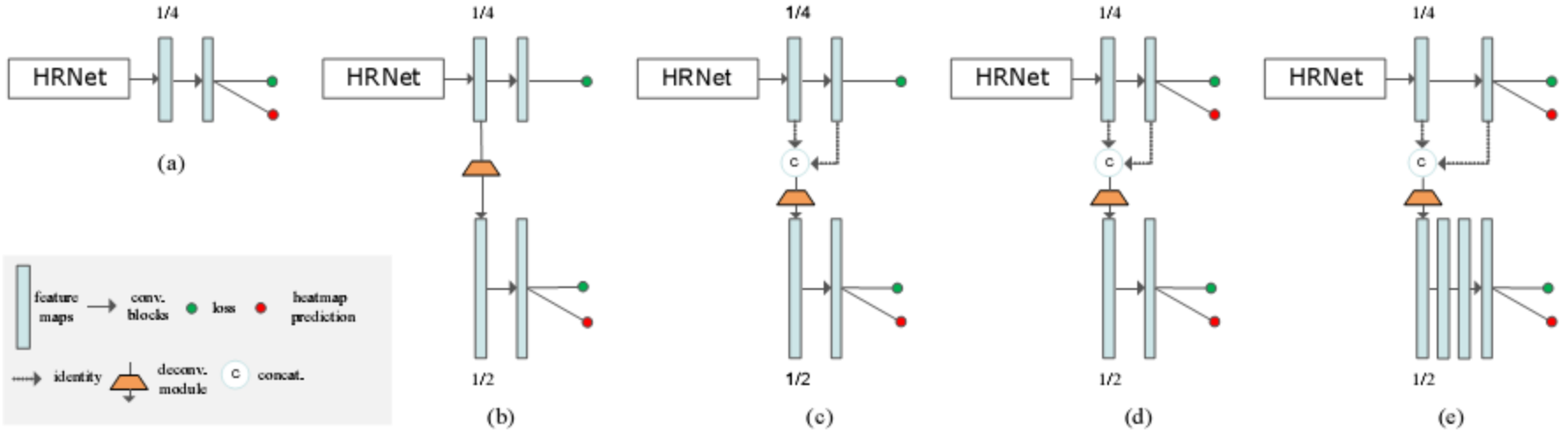
消融实验 higherhrnet详解
Ablation Experiments



	Network	w/ MRS	feature concat.	w/ heatmap aggregation	extra res. blocks	AP	AP^M	AP^L
(a)	HRNet					64.4	57.1	75.6
(b)	HigherHRNet	✓				66.0	60.7	74.2
(c)	HigherHRNet	✓	✓			66.3	60.8	74.0
(d)	HigherHRNet	✓	✓	✓		66.9	61.0	75.7
(e)	HigherHRNet	✓	✓	✓	✓	67.1	61.5	76.1

反卷积模块的效果。（a）和（b）对比说明了更大的特征图上预测关键点效果更好。从64.4提升到66.0，提升了1.6。这证明了作者的观点，对自下而上的姿势估计，更高分辨率的特征图对预测有重要影响。

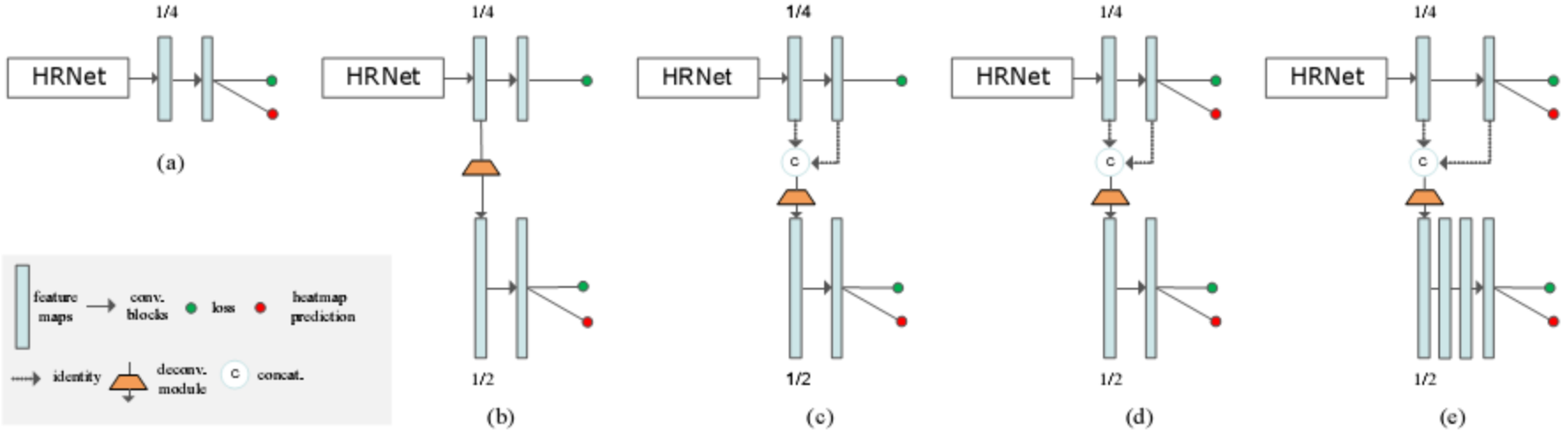
消融实验 higherhrnet详解
Ablation Experiments



	Network	w/ MRS	feature concat.	w/ heatmap aggregation	extra res. blocks	AP	AP^M	AP^L
(a)	HRNet					64.4	57.1	75.6
(b)	HigherHRNet	✓				66.0	60.7	74.2
(c)	HigherHRNet	✓	✓			66.3	60.8	74.0
(d)	HigherHRNet	✓	✓	✓		66.9	61.0	75.7
(e)	HigherHRNet	✓	✓	✓	✓	67.1	61.5	76.1

特征连接的效果。将特征图和热力图拼接输入反卷积，（c）的效果提升到了66.3。比较（a）和（c），中等大小的目标提升3.7，但是大目标下降了1.6。这证明了作者的观点，不同分辨率的特征图对不同尺寸的人敏感。

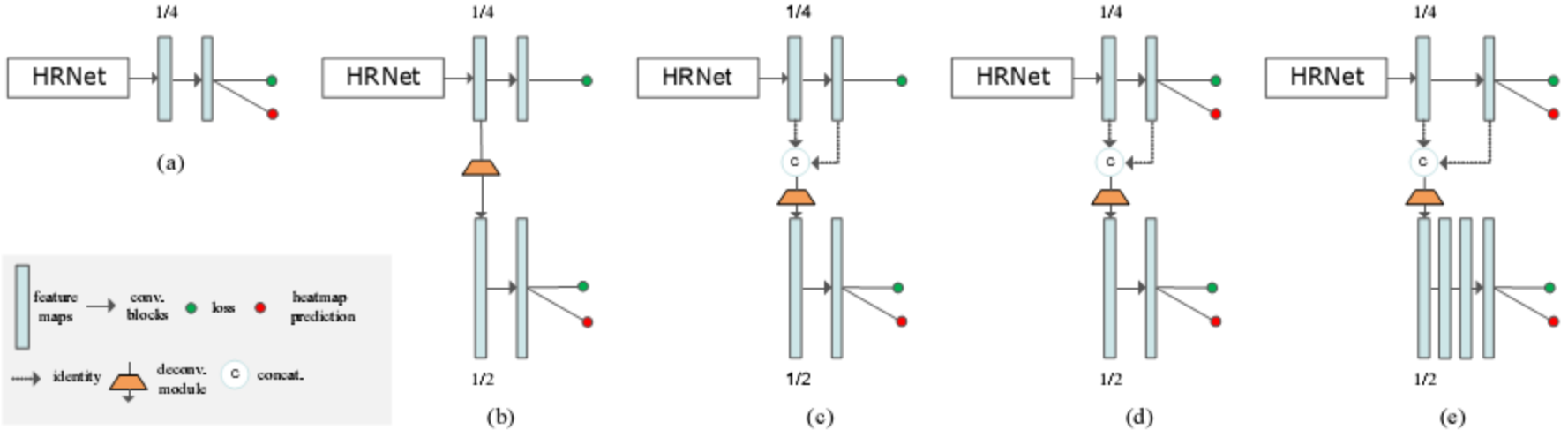
消融实验 higherhrnet详解
Ablation Experiments



	Network	w/ MRS	feature concat.	w/ heatmap aggregation	extra res. blocks	AP	AP^M	AP^L
(a)	HRNet					64.4	57.1	75.6
(b)	HigherHRNet	✓				66.0	60.7	74.2
(c)	HigherHRNet	✓	✓			66.3	60.8	74.0
(d)	HigherHRNet	✓	✓	✓		66.9	61.0	75.7
(e)	HigherHRNet	✓	✓	✓	✓	67.1	61.5	76.1

热力图融合的效果。（d）中根据热力图融合策略在推理时使用了所有分辨率的热力图，与（c）只使用最高分辨率的热力图比较，提升到了66.9。在大尺寸上与（c）比较提升了1.7，甚至与（a）比较稍微好一点。这证明预测使用热力图融合达到了尺寸感知的效果。

消融实验 higherhrnet详解
Ablation Experiments



	Network	w/ MRS	feature concat.	w/ heatmap aggregation	extra res. blocks	AP	AP^M	AP^L
(a)	HRNet					64.4	57.1	75.6
(b)	HigherHRNet	✓				66.0	60.7	74.2
(c)	HigherHRNet	✓	✓			66.3	60.8	74.0
(d)	HigherHRNet	✓	✓	✓		66.9	61.0	75.7
(e)	HigherHRNet	✓	✓	✓	✓	67.1	61.5	76.1

额外的残差块的效果。（e）在反卷积模块中增加了四个残差块，效果达到了最好的67.1。添加残差块进一步改善了特征图，在所有目标上均有提升。

消融实验 输入图片尺寸
Ablation Experiments

Network	Input Size	GFLOPs	AP
HRNet	256	9.7	43.5
HRNet	384	21.9	57.0
HRNet	512	38.9	64.4
HRNet	640	60.8	65.4
HRNet	768	87.5	63.3
HRNet	896	119.1	58.8
HRNet	1024	155.6	54.3
HigherHRNet	256	11.2	$52.3_{\uparrow 7.8}$
HigherHRNet	384	25.1	$63.4_{\uparrow 6.4}$
HigherHRNet	512	44.6	$67.1_{\uparrow 2.7}$
HigherHRNet	640	69.7	$67.4_{\uparrow 2.0}$
HigherHRNet	768	100.4	$64.7_{\uparrow 1.3}$
HigherHRNet	896	136.6	$61.0_{\uparrow 1.2}$
HigherHRNet	1024	178.4	$55.9_{\uparrow 1.6}$

Table 5. Ablation study of HigherHRNet with different input image size.

表5分析了输入尺寸的影响。随着尺寸减少，hrnet和higherhrnet差距在变大。这意味着分辨率下降对higherhrnet的影响更少。在低分辨率的姿势估计中生成更高分辨率的特征图是取得好效果的关键。

消融实验 输入图片尺寸
Ablation Experiments

Network	Input Size	GFLOPs	AP
HRNet	256	9.7	43.5
HRNet	384	21.9	57.0
HRNet	512	38.9	64.4
HRNet	640	60.8	65.4
HRNet	768	87.5	63.3
HRNet	896	119.1	58.8
HRNet	1024	155.6	54.3
HigherHRNet	256	11.2	$52.3_{\uparrow 7.8}$
HigherHRNet	384	25.1	$63.4_{\uparrow 6.4}$
HigherHRNet	512	44.6	$67.1_{\uparrow 2.7}$
HigherHRNet	640	69.7	$67.4_{\uparrow 2.0}$
HigherHRNet	768	100.4	$64.7_{\uparrow 1.3}$
HigherHRNet	896	136.6	$61.0_{\uparrow 1.2}$
HigherHRNet	1024	178.4	$55.9_{\uparrow 1.6}$

Table 5. Ablation study of HigherHRNet with different input image size.

因此当需要较小的输入分辨率和计算复杂度重要时，higherhrnet是一个好的选择。输入尺寸384的higherhrnet与输入尺寸512的hrnet的表现相当，但有13.8GFLOPs（相当于36%）被减少。

消融实验 使用更大尺寸训练
Ablation Experiments

Training size	AP	AP^M	AP^L
512	67.1	61.5	76.1
640	68.5	64.3	75.3
768	68.5	64.9	73.8

Table 6. Ablation study of HigherRNet with different training image size.

HigherHRNet	256	11.2	$52.3_{\uparrow 7.8}$
HigherHRNet	384	25.1	$63.4_{\uparrow 6.4}$
HigherHRNet	512	44.6	$67.1_{\uparrow 2.7}$
HigherHRNet	640	69.7	$67.4_{\uparrow 2.0}$
HigherHRNet	768	100.4	$64.7_{\uparrow 1.3}$
HigherHRNet	896	136.6	$61.0_{\uparrow 1.2}$
HigherHRNet	1024	178.4	$55.9_{\uparrow 1.6}$

在表5中发现使用512输入尺寸训练的higherhrnet在测试时使用640尺寸表现最佳。

一个自然的问题产生了，使用更大的输入尺寸能提升性能吗？

所以论文作者又训练了640和768尺寸。

消融实验 使用更大尺寸训练
Ablation Experiments

Training size	AP	AP^M	AP^L
512	67.1	61.5	76.1
640	68.5	64.3	75.3
768	68.5	64.9	73.8

Table 6. Ablation study of HigherRNet with different training image size.

三个模型都使用训练尺寸测试。

将训练尺寸提升到640，性能提升了1.4 ap。

主要提升在中等目标，大目标性能轻微下降。

当提升到768，总的ap不再变化，伴随着中等目标性能小提升和大目标性能严重下降。

消融实验 Larger backbone Ablation Experiments

Backbone	#Params	GFLOPs	AP	AP^M	AP^L
HRNet-W32	28.6	47.8	68.5	64.3	75.3
HRNet-W40	44.5	110.7	69.2	64.9	75.9
HRNet-W48	63.8	154.3	69.9	65.4	76.4

Table 7. Ablation study of HigherRNet with different backbone.

前面的实验都是使用hrnet-w32（1/4分辨率的特征图通道数32）作为主干。

现在使用hrnet-w40和hrnet-w48做主干对比。结果展示在表7。

使用更大的主干对所有目标能相应的提升性能。

problem

- higherhrnet网络输出两种分辨率特征图的作用？
- 与resnet相比，Hrnet预测热力图是否有优势，原因是什么？