AttentionMask: Attentive, Efficient Object Proposal Generation Focusing on Small Objects

Christian Wilms, Simone Frintrop

University of Hamburg
Department of Informatics

04 December 2018
ACCV 2018, Perth

Motivation - What is Object Proposal Generation?

Object Proposal Generation

Motivation - What is Object Proposal Generation?

Object Proposal Generation

- k **class-agnostic** object proposals generated
- No classification involved
- Localization with masks or boxes
- Proposals get an objectness score for ranking purpose
- Object proposal generation is usually a first step in object detection (reducing complexity)

Motivation - What is Object Proposal Generation?

Object Proposal Generation

Obj_1, Obj_2, Obj_3,...Obj_k

- k **class-agnostic** object proposals generated
- No classification involved
- Localization with masks or boxes
- Proposals get an objectness score for ranking purpose
- Object proposal generation is usually a first step in object detection (reducing complexity)

Motivation - Problems of Current Systems

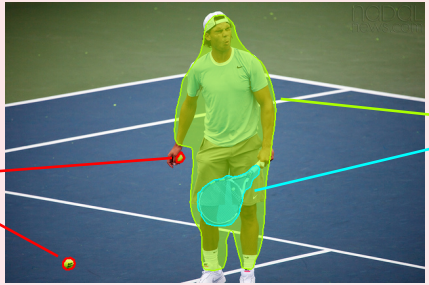
Problem 1: Small Objects

Problem 2: Inefficient Use of Resources

Motivation - Problems of Current Systems

Problem 1: Small Objects

- Small objects are regularly missed by state-of-the-art systems
- Main reason: Subsampling in CNNs for deeper features



Problem 2: Inefficient Use of Resources

Motivation - Problems of Current Systems

Problem 1: Small Objects

Problem 2: Inefficient Use of Resources

- Systems are either not really fast ($> 2s$ per image, SharpMaskZoom [1]) or...
- ... have a huge memory footprint (approx. 12GB on the GPU, FastMask [2]), preventing a more detailed analysis of the image

[1]: P. Pinheiro, T. Lin, R. Collobert and P. Dollàr, Learning to Refine Object Segments, in ECCV, 2016.
[2]: H. Hu, S. Lan, Y. Jiang, Z. Cao, F. Sha, FastMask: Segment Multi-scale Object Candidates in One Shot, in CVPR, 2017.

Motivation - Problems of Current Systems

Problem 1: Small Objects

Problem 2: Inefficient Use of Resources

Possible Solution:
Use the resources more efficiently and make use of the freed resources for better detecting small objects.

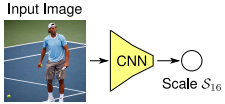
AttentionMask

Input Image

Idea

- Start from an existing system (FastMask [2])

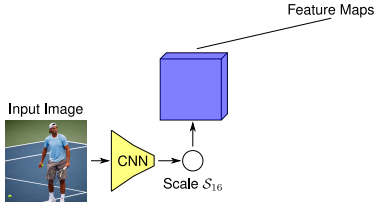
AttentionMask



Idea

- Start from an existing system (FastMask [2])

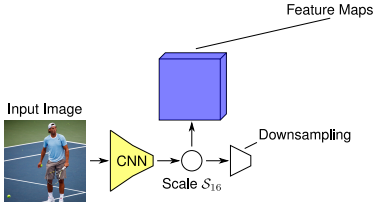
AttentionMask



Idea

- Start from an existing system (FastMask [2])

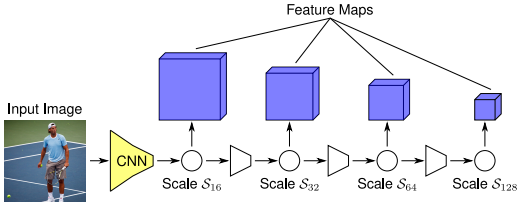
AttentionMask



Idea

- Start from an existing system (FastMask [2])

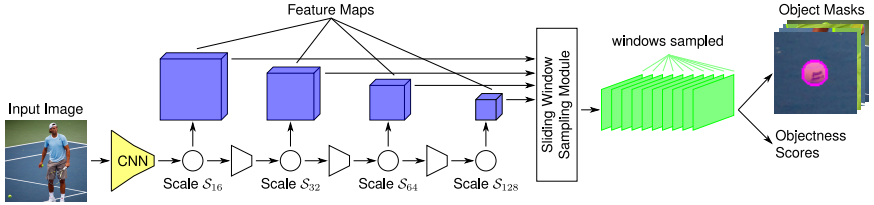
AttentionMask



Idea

- Start from an existing system (FastMask [2])

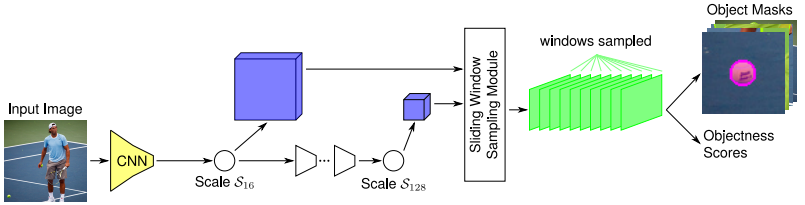
AttentionMask



Idea

- Start from an existing system (FastMask [2])

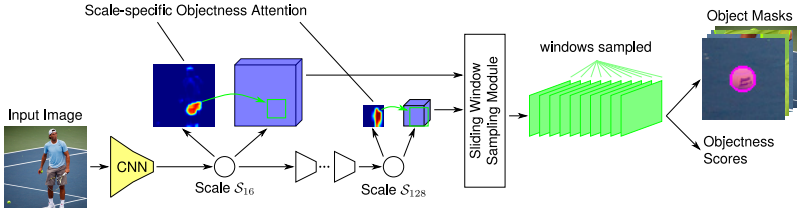
AttentionMask



Idea

- Start from an existing system (FastMask [2])

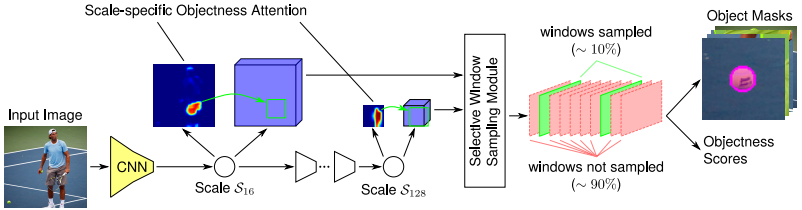
AttentionMask



Idea

- Start from an existing system (FastMask [2])
- Add attention modules at each scale and learn to focus on objects of relevant size

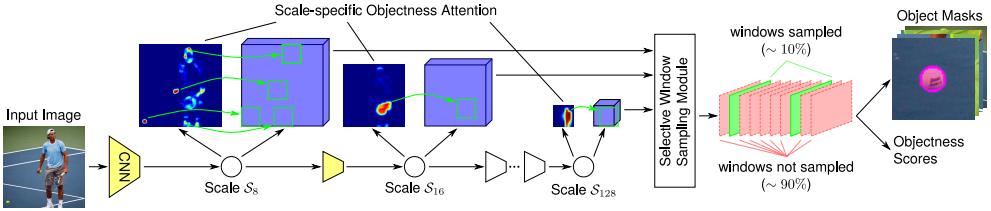
AttentionMask



Idea

- Start from an existing system (FastMask [2])
- Add attention modules at each scale and learn to focus on objects of relevant size
- Selectively sample windows in the feature maps given the attention

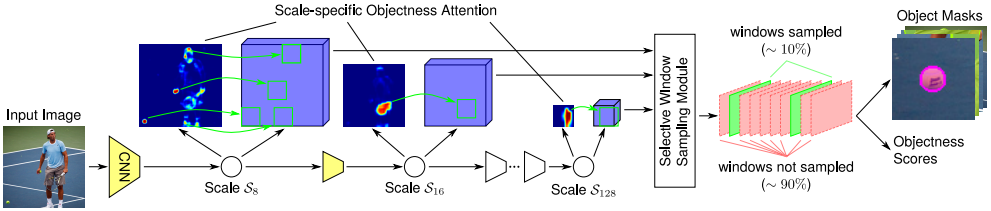
AttentionMask



Idea

- Start from an existing system (FastMask [2])
- Add attention modules at each scale and learn to focus on objects of relevant size
- Selectively sample windows in the feature maps given the attention
- **Lower memory consumption based on attention $\Rightarrow$ new scale ($\mathcal{S}_8$) for small objects possible**

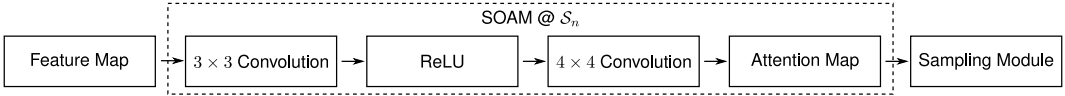
AttentionMask



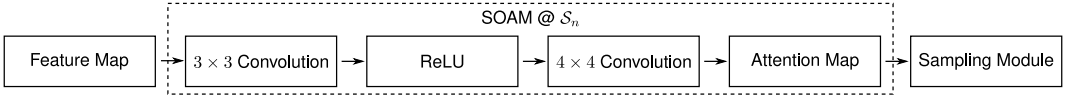
Idea

- Start from an existing system (FastMask [2])
- Add attention modules at each scale and learn to focus on objects of relevant size
- Selectively sample windows in the feature maps given the attention
- **Lower memory consumption based on attention $\Rightarrow$ new scale ($\mathcal{S}_8$) for small objects possible**
- System is trained end-to-end with multiple tasks

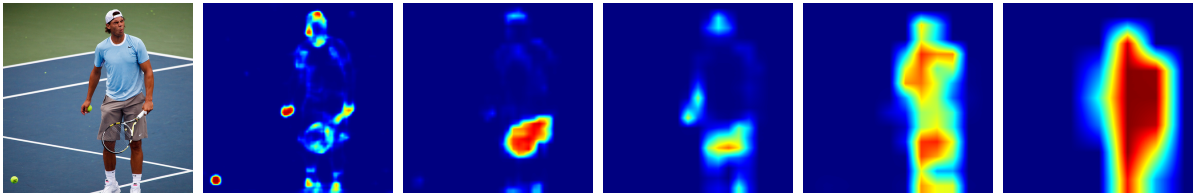
Scale-specific Objectness Attention Module (SOAM)



Scale-specific Objectness Attention Module (SOAM)

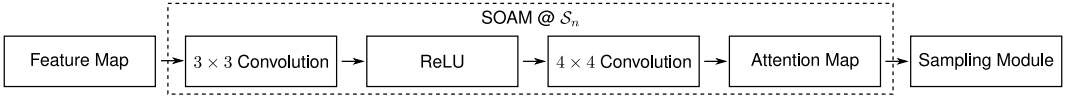


Attention maps for different scales:

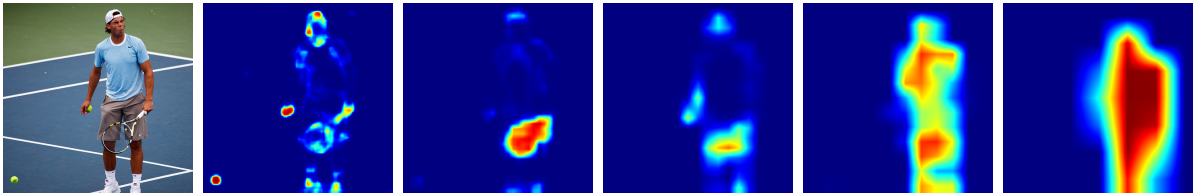


Input $\mathcal{S}_8$ $\mathcal{S}_{16}$ $\mathcal{S}_{32}$ $\mathcal{S}_{64}$ $\mathcal{S}_{128}$

Scale-specific Objectness Attention Module (SOAM)



Attention maps for different scales:



Input $\mathcal{S}_8$ $\mathcal{S}_{16}$ $\mathcal{S}_{32}$ $\mathcal{S}_{64}$ $\mathcal{S}_{128}$

Attention maps focus on objects of different scale!

Evaluation

- MS COCO dataset for training, validation and testing (80000/5000/5000)
- Average recall (AR) is used as evaluation measure
- AR measures how many GT objects are found and how well they are found
- AR correlates with the results of object detection using these proposals

Evaluation

- MS COCO dataset for training, validation and testing (80000/5000/5000)
- Average recall (AR) is used as evaluation measure
- AR measures how many GT objects are found and how well they are found
- AR correlates with the results of object detection using these proposals

Method	AR@10	AR@100	AR@1k	AR^S@100	AR^M@100	AR^L@100	Time
MCG [3]	0.077	0.186	0.299	-	-	-	45s
SharpMaskZoom [1]	0.156	0.304	0.401	0.099	0.412	0.495	2.02s
FastMask [2]	0.169	0.313	0.406	0.106	0.406	0.517	0.33s
AttentionMask	**0.180**	**0.349**	**0.444**	**0.162**	**0.421**	**0.560**	**0.22s**

[3]: J. Pont-Tuset, P. Arbelaez, J. Barron, F. Marques and J. Malik, Multiscale combinatorial grouping for image segmentation and object proposal generation, in TPAMI, 2017.

Evaluation

- MS COCO dataset for training, validation and testing (80000/5000/5000)
- Average recall (AR) is used as evaluation measure
- AR measures how many GT objects are found and how well they are found
- AR correlates with the results of object detection using these proposals

Method	AR@10	AR@100	AR@1k	AR^S@100	AR^M@100	AR^L@100	Time
MCG [3]	0.077	0.186	0.299	-	-	-	45s
SharpMaskZoom [1]	0.156	0.304	0.401	0.099	0.412	0.495	2.02s
FastMask [2]	0.169	0.313	0.406	0.106	0.406	0.517	0.33s
AttentionMask	**0.180**	**0.349**	**0.444**	**0.162**	**0.421**	**0.560**	**0.22s**

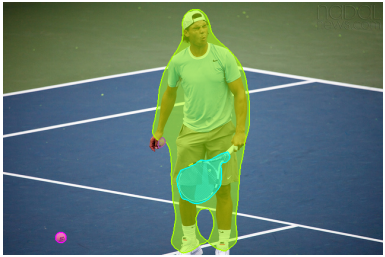
AttentionMask beats all state-of-the-art methods across all categories including runtime.

[3]: J. Pont-Tuset, P. Arbelaez, J. Barron, F. Marques and J. Malik, Multiscale combinatorial grouping for image segmentation and object proposal generation, in TPAMI, 2017.

Qualitative Results - I

SharpMaskZoom [1]

FastMask [2]



AttentionMask

Red non-filled shapes denote missed objects. Colored filled shapes denote found objects.

Qualitative Results - II

SharpMaskZoom [1]

FastMask [2]

AttentionMask

Red non-filled shapes denote missed objects. Colored filled shapes denote found objects.

Conclusion

- Small object are not well detected by state-of-the-art systems
- Idea: More efficient use of resources to put more effort in detecting small objects
- Introduction of SOAMs (attention) to focus processing
- Resources are now free for better detecting small objects
- Results prove efficiency and effectiveness of the approach

Conclusion

- Small object are not well detected by state-of-the-art systems
- Idea: More efficient use of resources to put more effort in detecting small objects
- Introduction of SOAMs (attention) to focus processing
- Resources are now free for better detecting small objects
- Results prove efficiency and effectiveness of the approach

Attention is a useful tool to focus processing in deep networks on important parts and make systems more efficient (and effective)!

Thank you for your attention!

AttentionMask: Attentive, Efficient Object Proposal Generation Focusing on Small Objects

For more information check out our webpage:
`https://www.inf.uni-hamburg.de/en/inst/ab/cv/people/wilms/attentionMask.html`

References I

- Pinheiro, P.O., Lin, T.Y., Collobert, R., Dollár, P.:
  Learning to refine object segments.
  In: ECCV. (2016)
- Hu, H., Lan, S., Jiang, Y., Cao, Z., Sha, F.:
  FastMask: Segment multi-scale object candidates in one shot.
  In: CVPR. (2017)
- Pont-Tuset, J., Arbelaez, P., Barron, J.T., Marques, F., Malik, J.:
  Multiscale combinatorial grouping for image segmentation and object proposal generation.
  TPAMI **39** (2017)