CLASS AGNOSTIC OBJECT SEGMENTATION with FEW-SHOT WEAKLY SUPERVISED GUIDANCE

Mennatullah Siam*, Naren Doraiswamy*,
Boris Oreshkin*, Hengshuai Yao, Martin Jagersand

Introduction

Few-Shot Learning impact on Developing Countries

Developing countries suffer from **limited computational resources and labelled datasets**.

Potential Applications: aerial images segmentation - perception for robot manipulation

Few-Shot Segmentation with Image-level Labels

Can use publicly available **web data**.

Literature mainly focused on **pixel level labels** and **bounding boxes,** with only one recent approach (**Raza et. al.[1]**) on **image-leve**l.

[1] Hasnain Raza, Mahdyar Ravanbakhsh, Tassilo Klein, and Moin Nabi. Weakly supervised one shot segmentation. In Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 0–0, 2019

Class Agnostic Segmentation with Few-shot Guidance

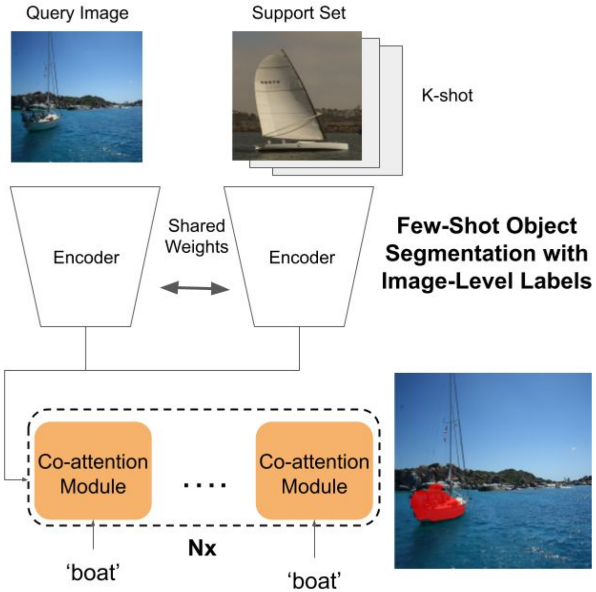
Setup: Following Shaban et al. setup **1-way k-shot** segmentation,

Where goal is to segment 1 class against background

Using k images as few training data

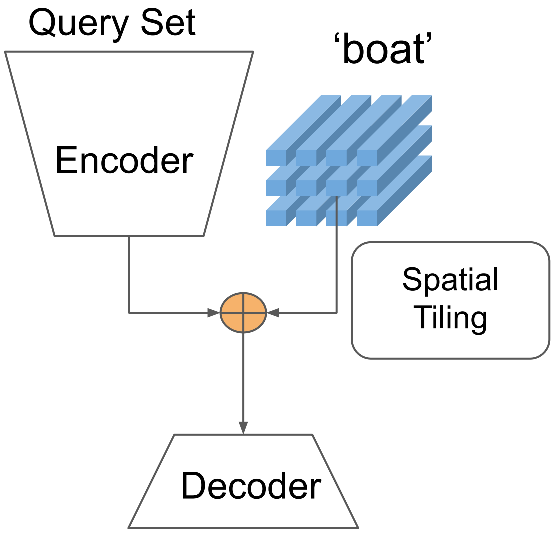
Contributions:

- Investigate co-attention mechanisms and conditioning on semantic features in learning **few-shot segmentation with image-level supervision**.
- **Temporal Object segmentation for few-shot learning** novel setup.
- **Comparative study** of different variants.

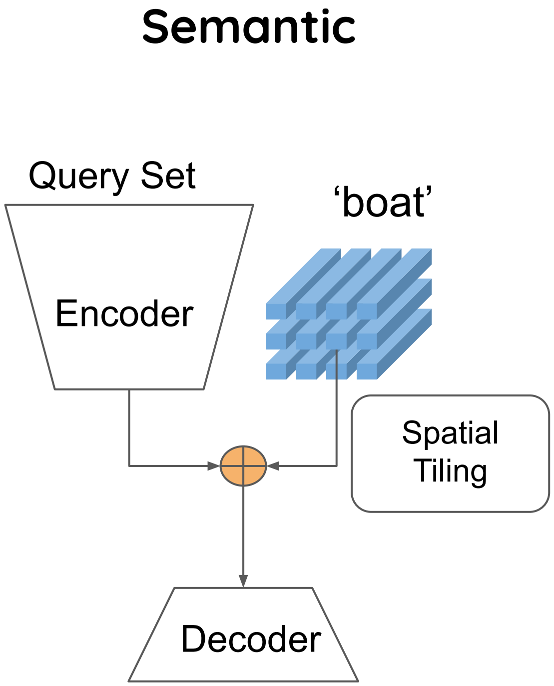


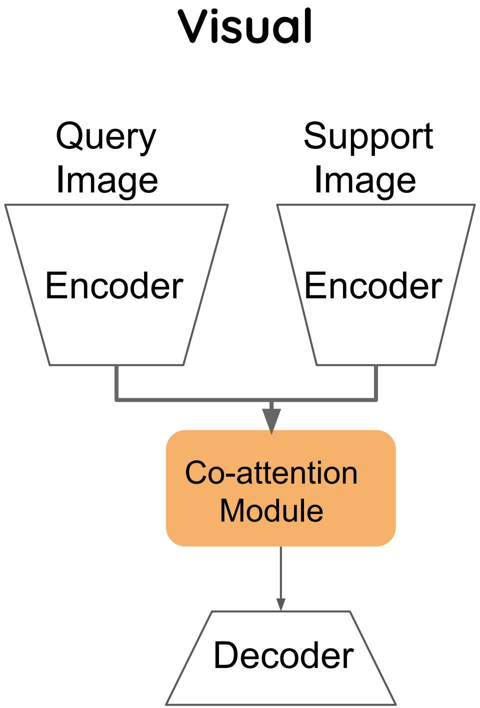
Class Agnostic Segmentation with Few-shot Guidance

Semantic

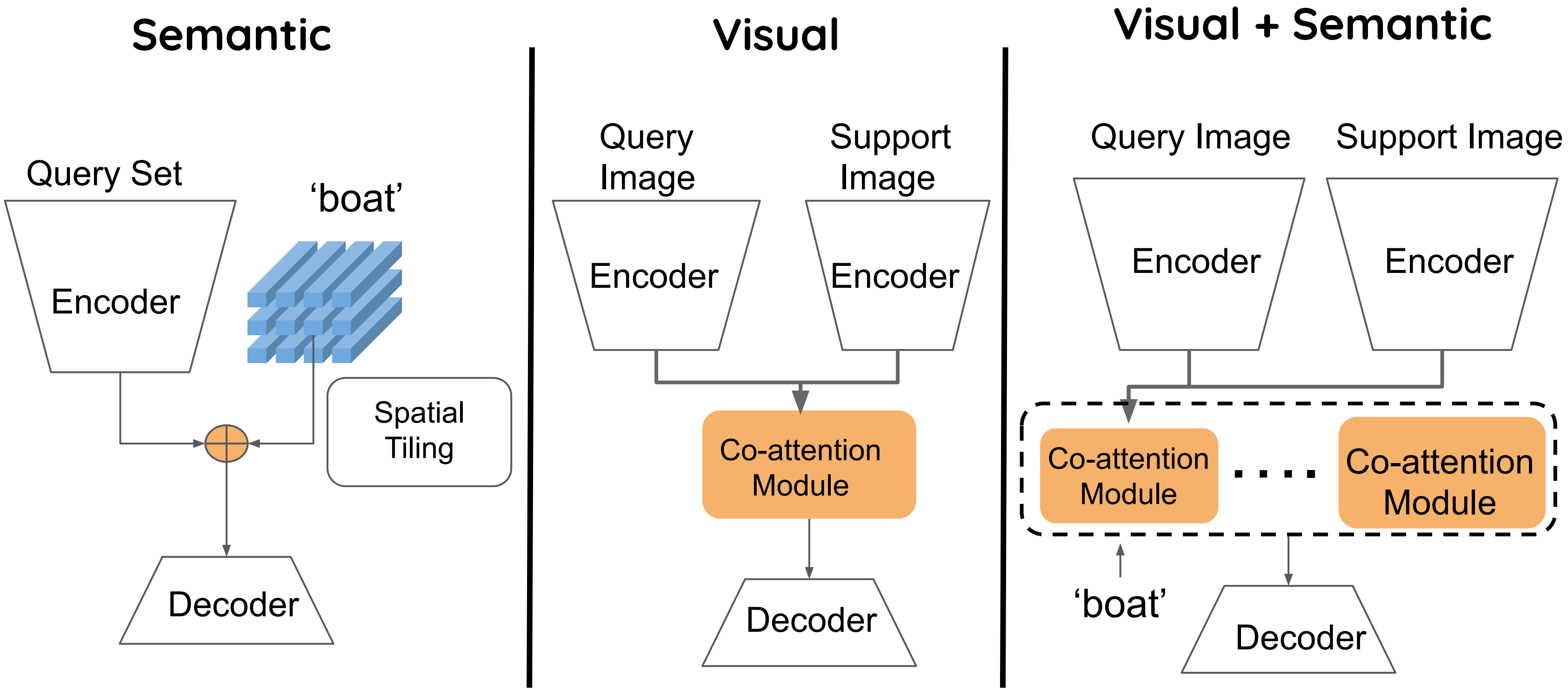


Class Agnostic Segmentation with Few-shot Guidance





Class Agnostic Segmentation with Few-shot Guidance



Class Agnostic Segmentation with Few-shot Guidance

Pixel-Level Affinity

$$S = \tilde{V_s}^T W_{co} \tilde{V_q}$$

$$S^c = \text{softmax}(S)$$

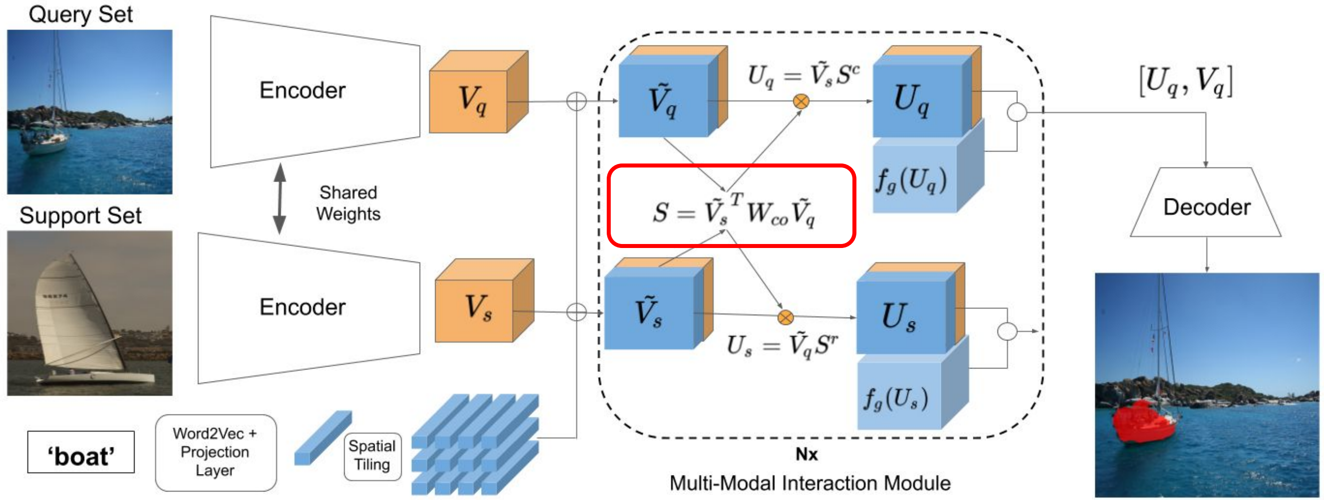
Attention Summaries

$$U_q = \tilde{V}_s S^c$$

Gated Attention

$$f_g(U_q) = \sigma(W_g * U_q + b_g)$$

$$U_q = f_g(U_q) \circ U_q$$



Few-Shot Object Segmentation with Image-Level Supervision

Class Agnostic Segmentation with Few-shot Guidance

Pixel-Level Affinity

$$S = \tilde{V_s}^T W_{co} \tilde{V_q}$$

$$S^c = \mathrm{softmax}(S)$$

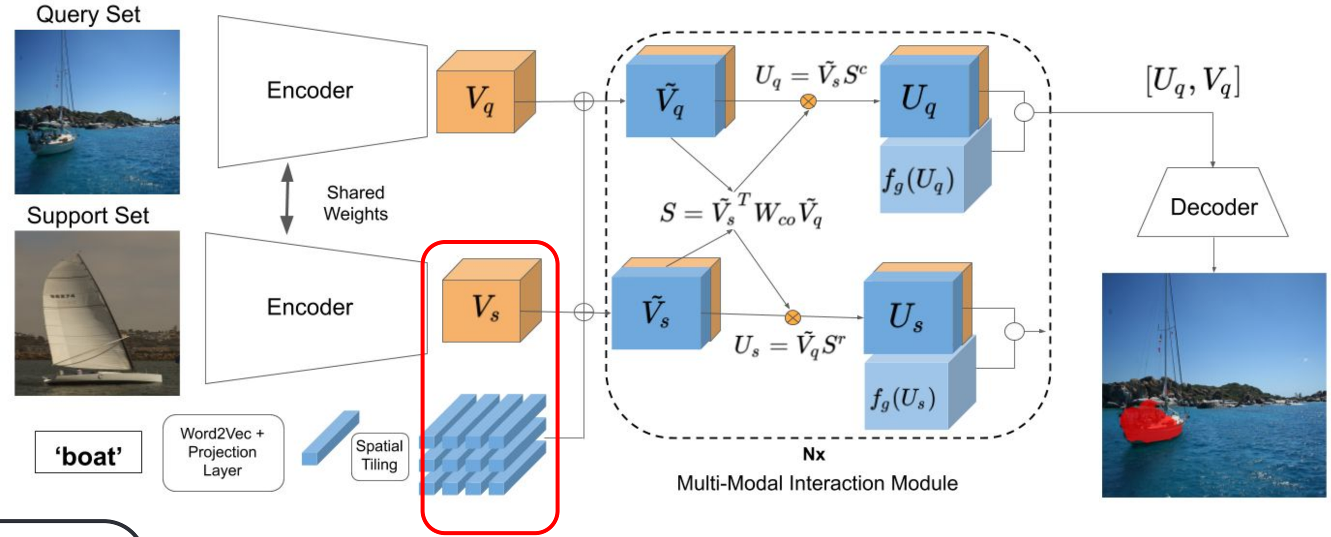
Attention Summaries

$$U_q = \tilde{V_s} S^c$$

Gated Attention

$$f_g(U_q) = \sigma(W_g * U_q + b_g)$$

$$U_q = f_g(U_q) \circ U_q$$



Few-Shot Object Segmentation with Image-Level Supervision

Class Agnostic Segmentation with Few-shot Guidance

Pixel-Level Affinity

$$S = \tilde{V_s}^T W_{co} \tilde{V_q}$$

$$S^c = \mathrm{softmax}(S)$$

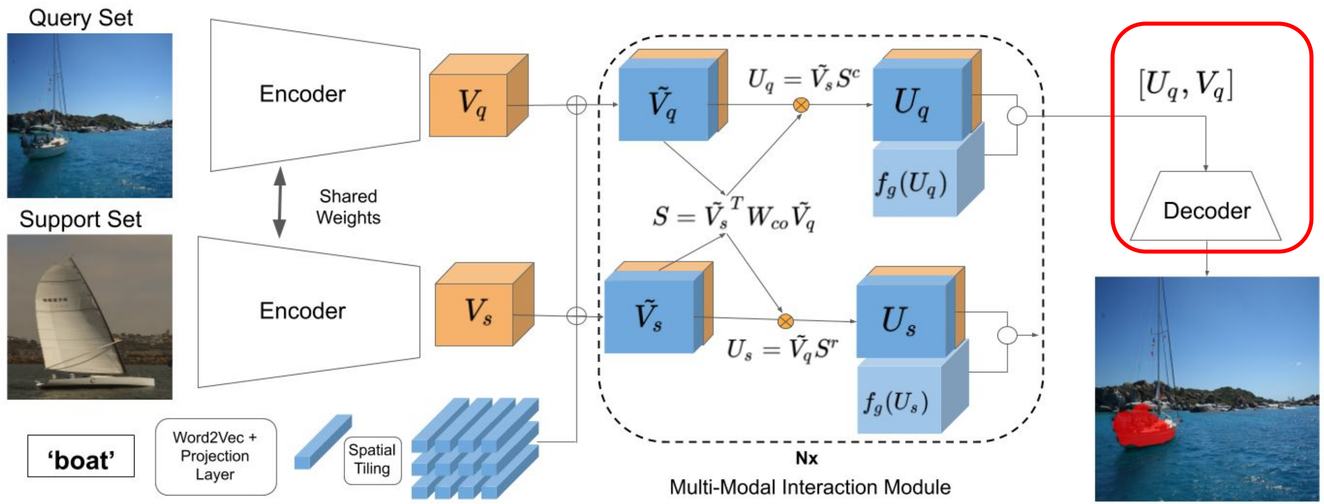
Attention Summaries

$$U_q = \tilde{V_s} S^c$$

Gated Attention

$$f_g(U_q) = \sigma(W_g * U_q + b_g)$$

$$U_q = f_g(U_q) \circ U_q$$



Few-Shot Object Segmentation with Image-Level Supervision

TOSFL: Temporal Object Segmentation for Few-shot Learning

Problem: Binary segmentation of Query Images from a video sequence based on provided support set image-label pair, label is only **image-level label.**

Setup: Instance-level / Category-level, query set is sampled from a video sequence

Instance Level

Category Level

Motivation:

1) Pixels that move together belong to the same object

2) Predicted Masked Embeddings are expected to be temporally consistent.

Experimental Results

Ablation Studies:

1) Pascal-5i

Coattention	Semantic	mIoU
✗	✗	42.7
✓	✗	44.6
✓	✓	**51.0**

2) Youtube-VOS

Coattention	Semantic + Visual	mIoU
✗	✓	42.3
✓	✓	**43.7**

Experimental Results

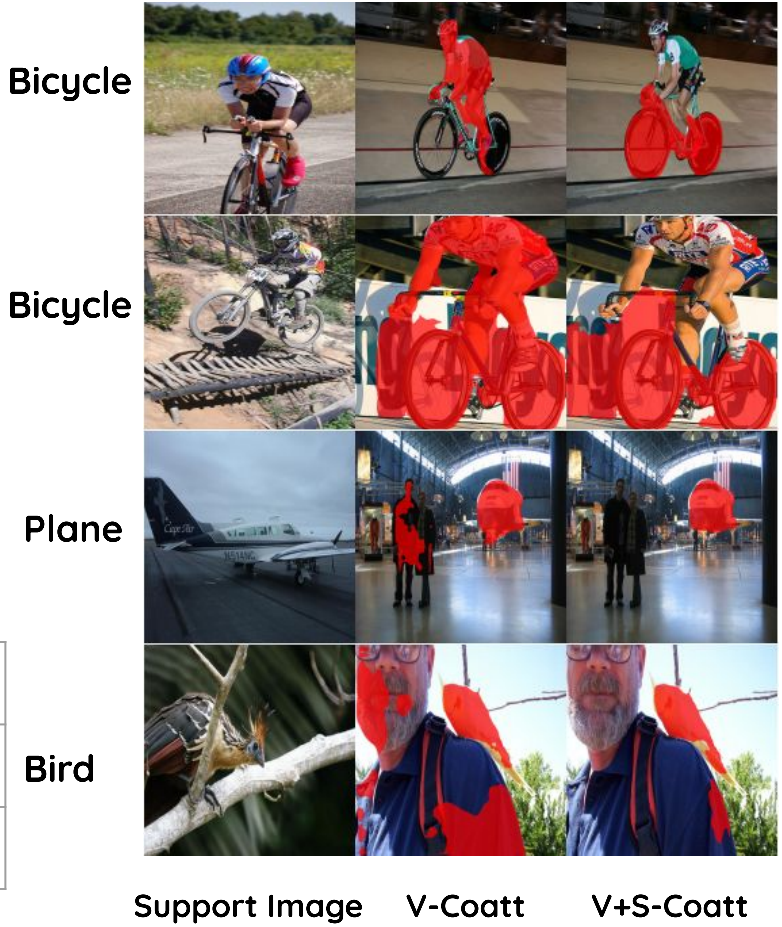
Ablation Studies:

1) Pascal-5i

Coattention	Semantic	mIoU
✗	✗	42.7
✓	✗	44.6
✓	✓	**51.0**

2) Youtube-VOS

Coattention	Semantic + Visual	mIoU
✗	✓	42.3
✓	✓	**43.7**



Experimental Results

Variants:

1) **Pascal-5i**

Method	1-shot	5-shot
V-CoAtt	44.4 ± 0.3	49.1 ± 0.3
S-Cond	**51.2 ± 0.6**	51.4 ± 0.3
V+S-Coatt	50.5 ± 0.7	**51.7 ± 0.07**

2) **Youtube-VOS**

Method	Category-Level	Instance Level
V-CoAtt	36.1	38.0 ± 0.7
S-Cond	**37.7**	41.7 ± 0.7
V+S-Coatt	37.6	**43.8 ± 0.5**

Experimental Results

Comparison to SOA

Method	Type	1-shot		5-shot
		mIoU	bIoU	mIoU
FG-BG	P	-	55.1	-
OSLSM (Shaban et. al. 2017)	P	40.8	-	43.9
CoFCN (Rakelly et. al. 2018)	P	41.1	60.1	41.4
PLSeg (Dong et. al. 2018)	P	-	61.2	-
AMP (Siam et. al. 2019)	P	43.4	62.2	46.9
PGNet (Zang et. al. 2019)	P	**56.0**	**69.9**	**58.5**
Ours	IL	**50.5**	**64.1**	**51.7**

Experimental Results

Comparison to SOA

Method	Type	1-shot		5-shot
		mIoU	bIoU	mIoU
PANet (Wang et. al 2019)	P	48.1	66.5	55.7
CANet (Zang et. al. 2019)	P	55.4	66.2	57.1
CANet (Zang et. al. 2019)	BB	**52.0**	-	-
PANet (Wang et. al 2019)	BB	**45.1**	-	**52.8**
(Raza et. al. 2019)	IL	-	**58.7**	-
Ours - V1	IL	**53.5**	**65.6**	**-**
Ours - V2	IL	**50.5**	**64.1**	**51.7**

Semantic features + Co-attention

Few-shot Image-level Guidance

Novel TOSFL Setup