GROUND-TO-AERIAL IMAGE MATCHING

TWO DIFFERENT APPROACHES

AUTHORS: MAFFONGELLI MATTIA, RAGO SALVATORE MICHELE

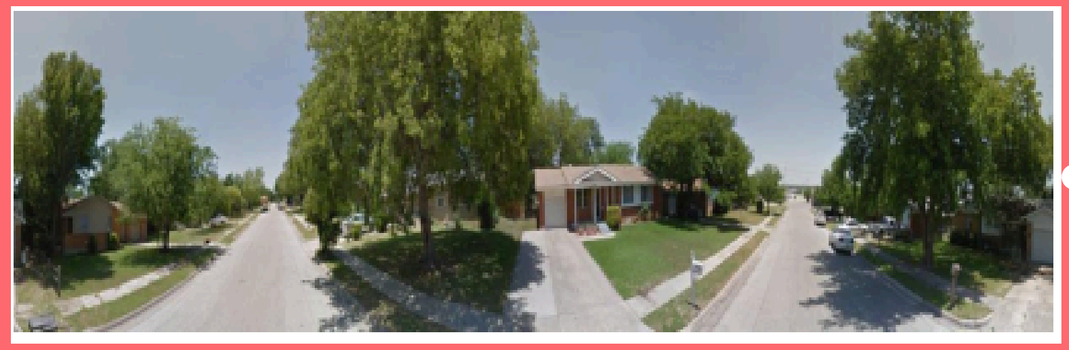
SAPIENZA
UNIVERSITÀ DI ROMA

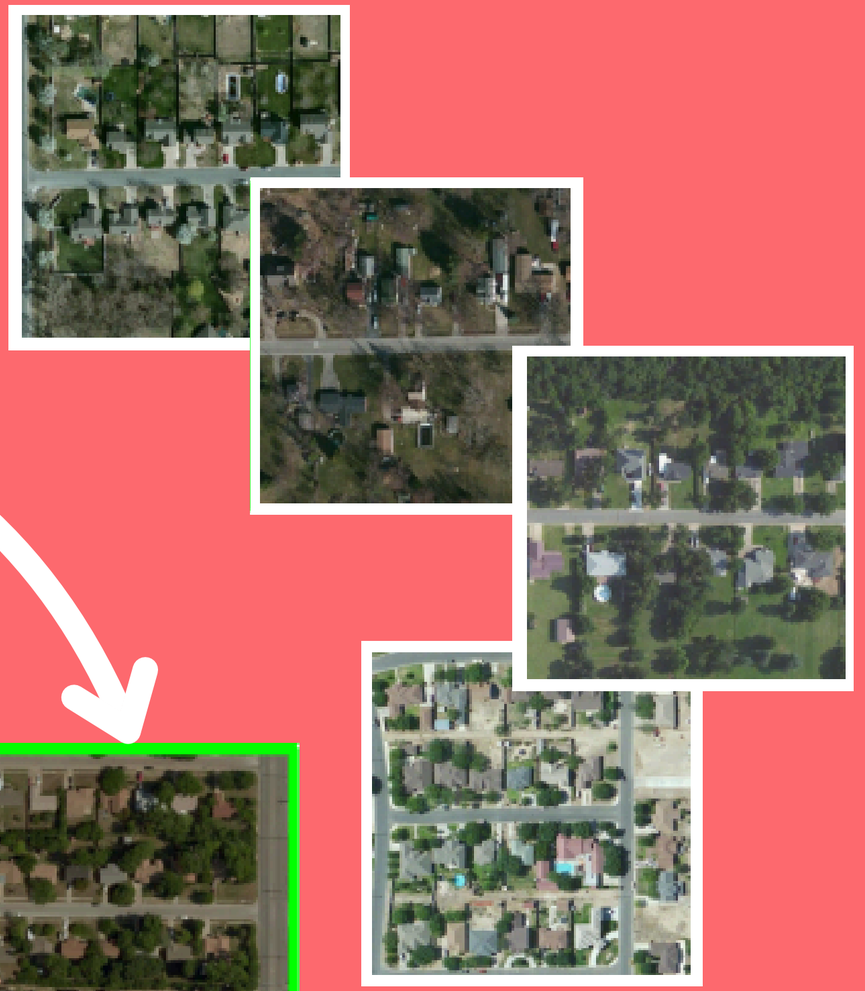
OUTLINE

- Introduction
- Related Works
- Implementation details
- Datasets and Metrics
- Experimental results
- Conclusion

SAPIENZA
UNIVERSITÀ DI ROMA

WHAT IS GROUND-TO-AERIAL MATCHING?





Main difficulties

- different alignment
- different PoV
- different image quality
- limited ground FoV

SAPIENZA
UNIVERSITÀ DI ROMA

RELATED WORKS

PRE-AI

Geometric constraints

3D reconstruction

POST-AI

GANs

ViTs

Siamese-like Networks

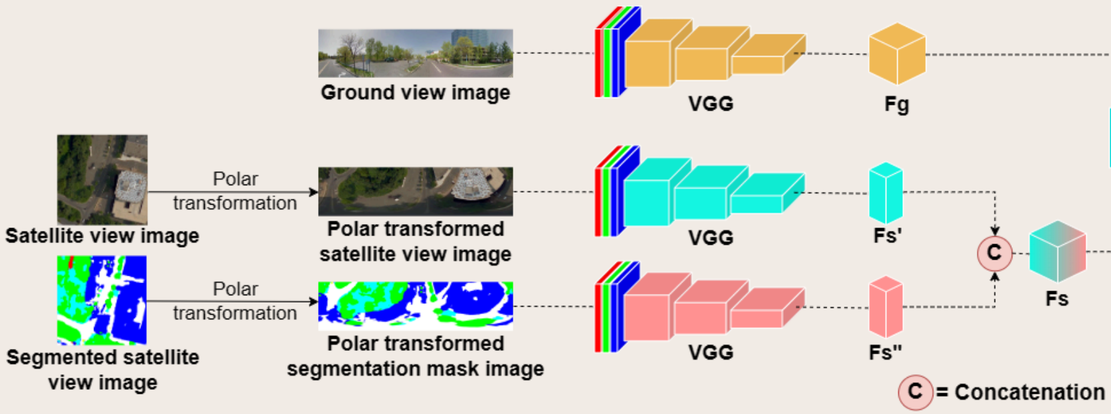
A SEMANTIC SEGMENTATION-GUIDED APPROACH FOR GROUND-TO-AERIAL IMAGE MATCHING (Pro et al. 2024)

WHAT'S THE INNOVATION?

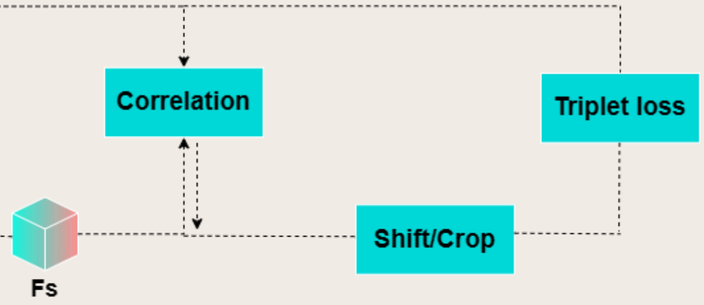
The primary innovation of this method lies in integrating satellite images with their corresponding segmentation masks. This approach ensures that the model can effectively extract relevant features by focusing specifically on significant parts of the images. So, the proposed Semantic Segmentation-Guided Approach (SAN), demonstrates enhanced performance within the unlabelled CVUSA dataset.

BUT, THE STRUCTURE OF THE SAN MODEL?

This architecture consists of three parallel VGG branches, useful for the features extraction: one for processing the ground-view image, one for the aerial image, and one for the semantic segmentation mask of the aerial image. In particular:



SAPIENZA UNIVERSITÀ DI ROMA



tenation

A SEMANTIC SEGMENTATION-GUIDED APPROACH FOR GROUND-TO-AERIAL IMAGE MATCHING

SO, THE FIRST STEP?

Since we have explained the structure, the first step of our work involves the segmentation of the satellite images. Let's see how this process works!

HOW IT WORKS?

SAPIENZA
UNIVERSITÀ DI ROMA

(UNLABELED) SEMANTIC SEGMENTATION

The semantic segmentation is the task of classifying each pixel in an image into a predefined set of categories or classes. Instead of relying on low-level features like keypoints and descriptors, which may not capture higher-level semantic information, by incorporating semantic segmentation, in this work, the matching algorithm gains a deeper understanding of the scene, having richer information of the context.

However, in the dataset used for training (CVUSA), there are no annotated aerial images (thus, with labels for roads, trees, and other objects). Therefore, It's proposed the NEOS method.

NEOS KEY POINTS

01. Labelled and unlabelled input for SegFormer

02. Three kinds of Loss functions

03. Data augmentation for more robustness

FURTHERMORE… POLAR TRANSFORMATION!

Another crucial point in preparing the input data for our project is the pre-processing step, involving polar transformation.

WHAT'S POLAR TRANSFORMATION?

Polar transformation is a technique used to mitigate scale and perspective differences between ground and aerial images. It helps the alignment part because the images would otherwise be distorted or mismatched due to variations in viewpoint. This kind of transformation is applied, following:

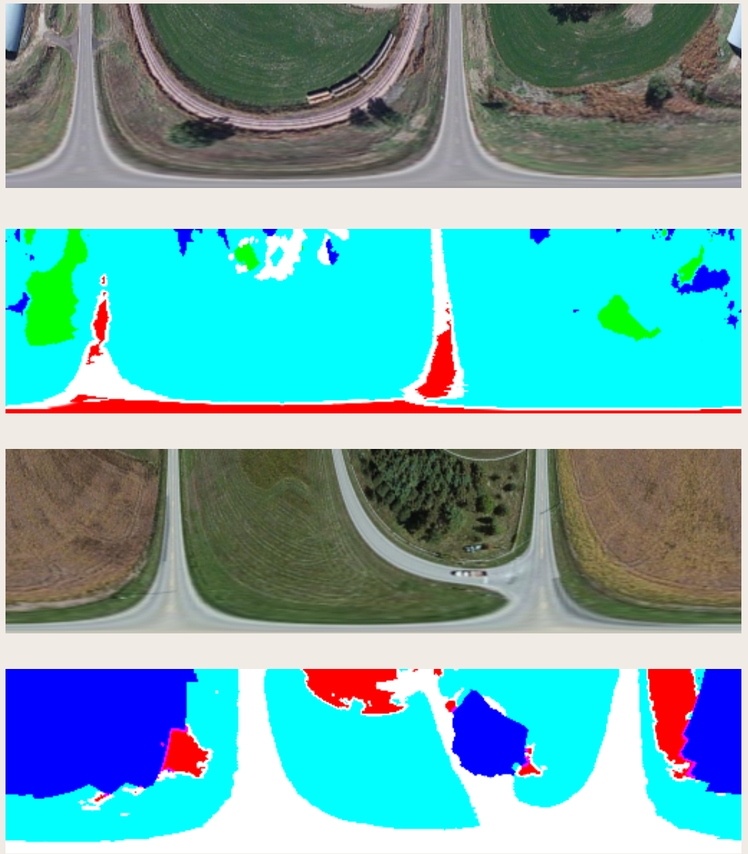
$$x_i^s = \frac{D_s}{2} - \frac{D_s}{2}\frac{(H_v - x_i^t)}{H_v}\cos(\frac{2\pi}{W_v}y_i^t) \quad (1)$$

$$y_i^s = \frac{D_s}{2} + \frac{D_s}{2}\frac{(H_v - x_i^t)}{H_v}\sin(\frac{2\pi}{W_v}y_i^t) \quad (2)$$

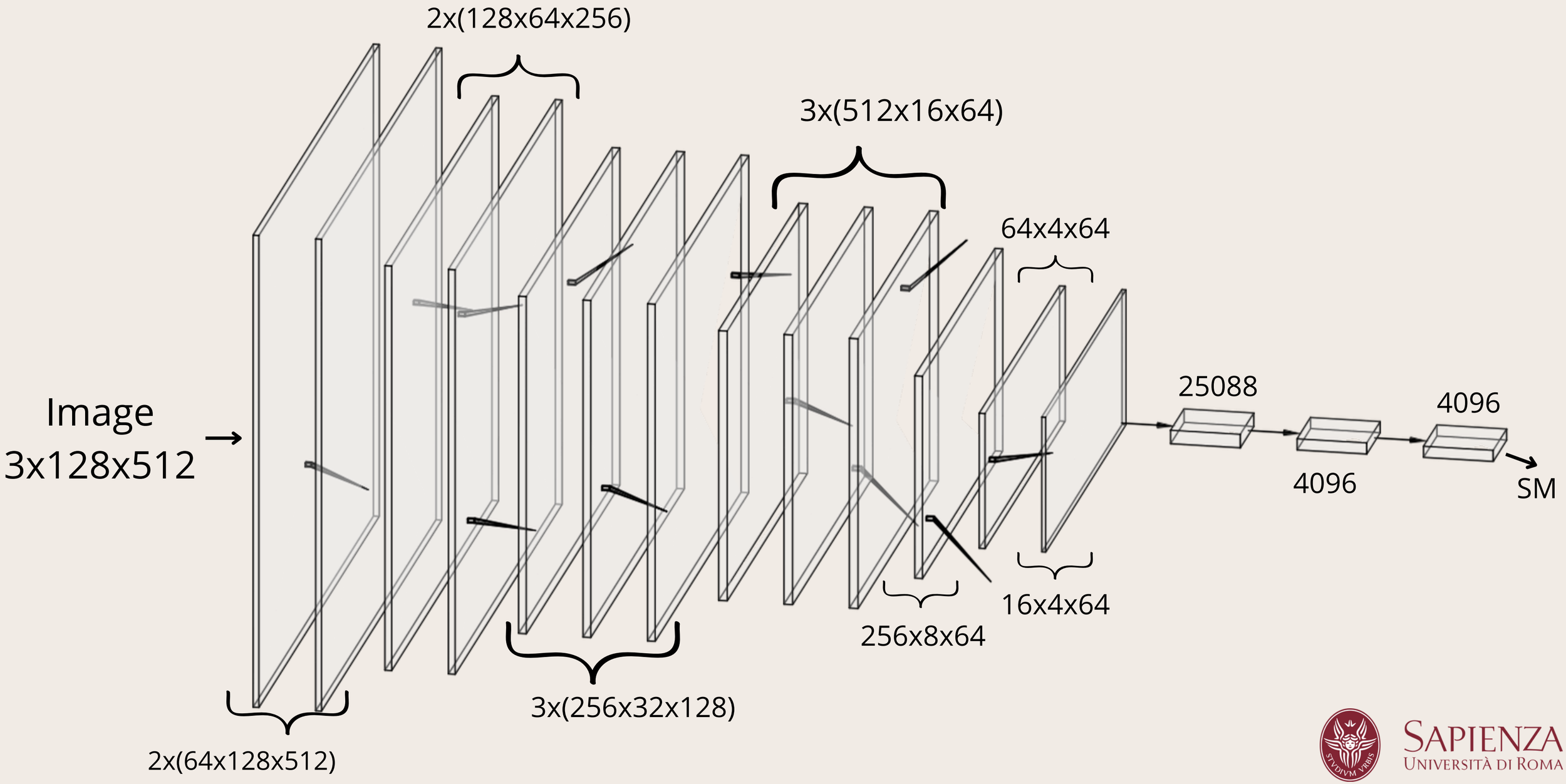
With: Ds x Ds is the size of an aerial image, and Hv × Wv denotes the target size of the polar transformation.

N.B.: Obviously, the images are later normalized and standardized!!!

RESULTS?

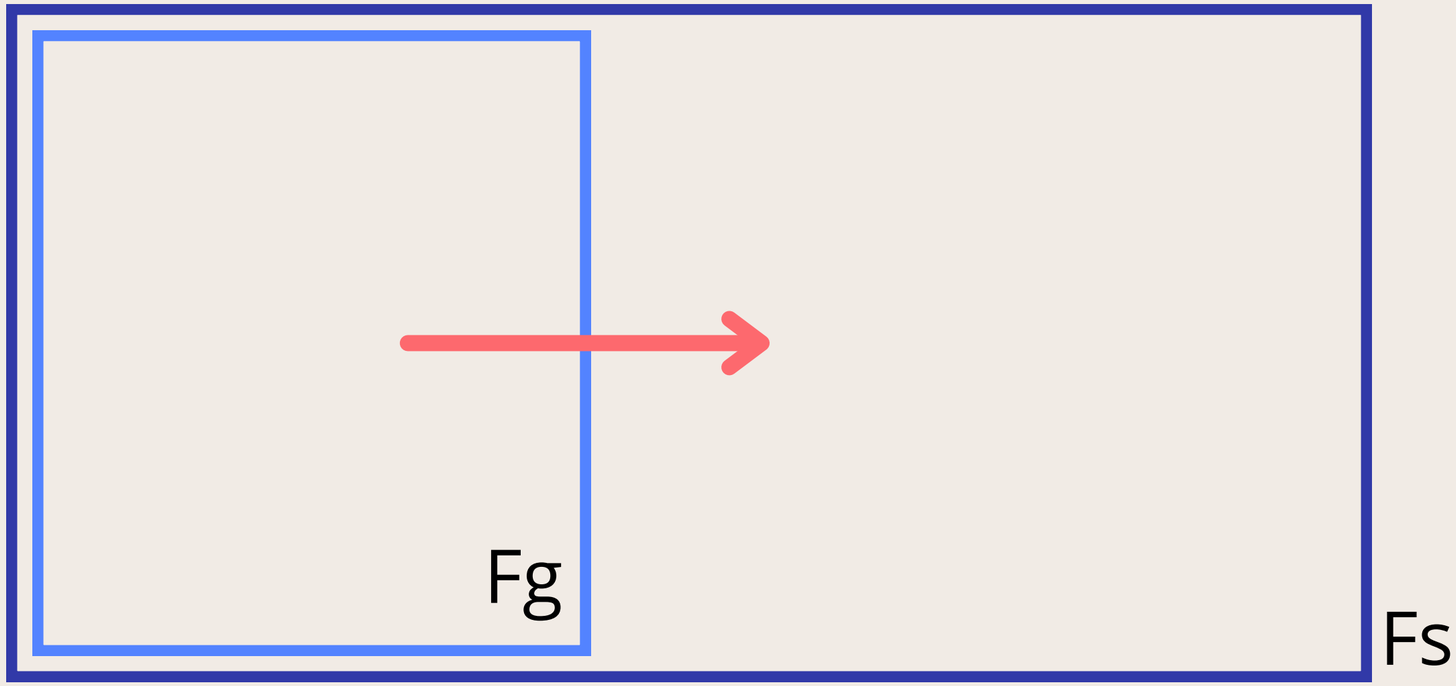


FEATURE EXTRACTOR (VGG16)



SAPIENZA
UNIVERSITÀ DI ROMA

CORRELATION



$$[F_s \star F_g](i) = \sum_{c=1}^{C} \sum_{h=1}^{H} \sum_{w=1}^{W_v} F_s(h, (i+w)\%W_s, c) \cdot F_g(h, w, c)$$

TRAINING

TRAINING PARAMETERS

DATASET CVUSA

EPOCHS 30

BATCH S. 8

L.R 10e-5

SAPIENZA
UNIVERSITÀ DI ROMA

EVALUTAION

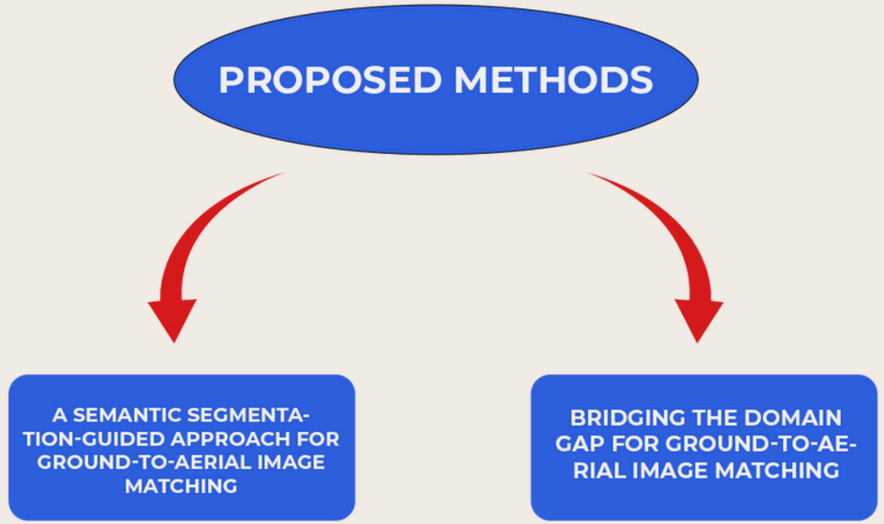
TOP-K METRIC FOR DIFFERENT FoV

Method	FoV 360°				FoV 180°			
	r@1	r@5	r@10	r@1%	r@1	r@5	r@10	r@1%
SAN (*our*)	**77,07%**	**92,14%**	**95,62%**	**97,97%**	**48,49%**	**75,53%**	**84,06%**	**91,24%**

FoV 90°				FoV 70°			
r@1	r@5	r@10	r@1%	r@1	r@5	r@10	r@1%
6,23%	**16,43%**	**24,20%**	**37,07%**	**3,02%**	**9,75%**	**15,44%**	**25,82%**

SAPIENZA
UNIVERSITÀ DI ROMA

NOW LET'S TALK ABOUT THE SECOND PROPOSED METHOD



WHAT ARE THE DIFFERENCES?

FIRST METHOD:

- The semantic segmentation-guided approach uses high-level semantic information to guide the matching process by focusing on meaningful parts of the images.
- Semantic segmentation helps in extracting meaningful features that are invariant to perspective changes.

SECOND METHOD:

- Domain adaptation’s approach focuses on reducing the domain gap between ground and aerial images (the differences in visual appearance, scale and perspective).
- It learns representations that are invariant to the domain (ground vs. aerial) so that features extracted from both domains can be directly compared.
- GAN for the synthesized aerial images, and two main methods: Joint Feature Learning and Features Fusion.

SAPIENZA
UNIVERSITÀ DI ROMA

HOW PRODUCE SYNTHESIZED IMAGES?

PAPER APPROACH:

GAN: X-FORK MODEL

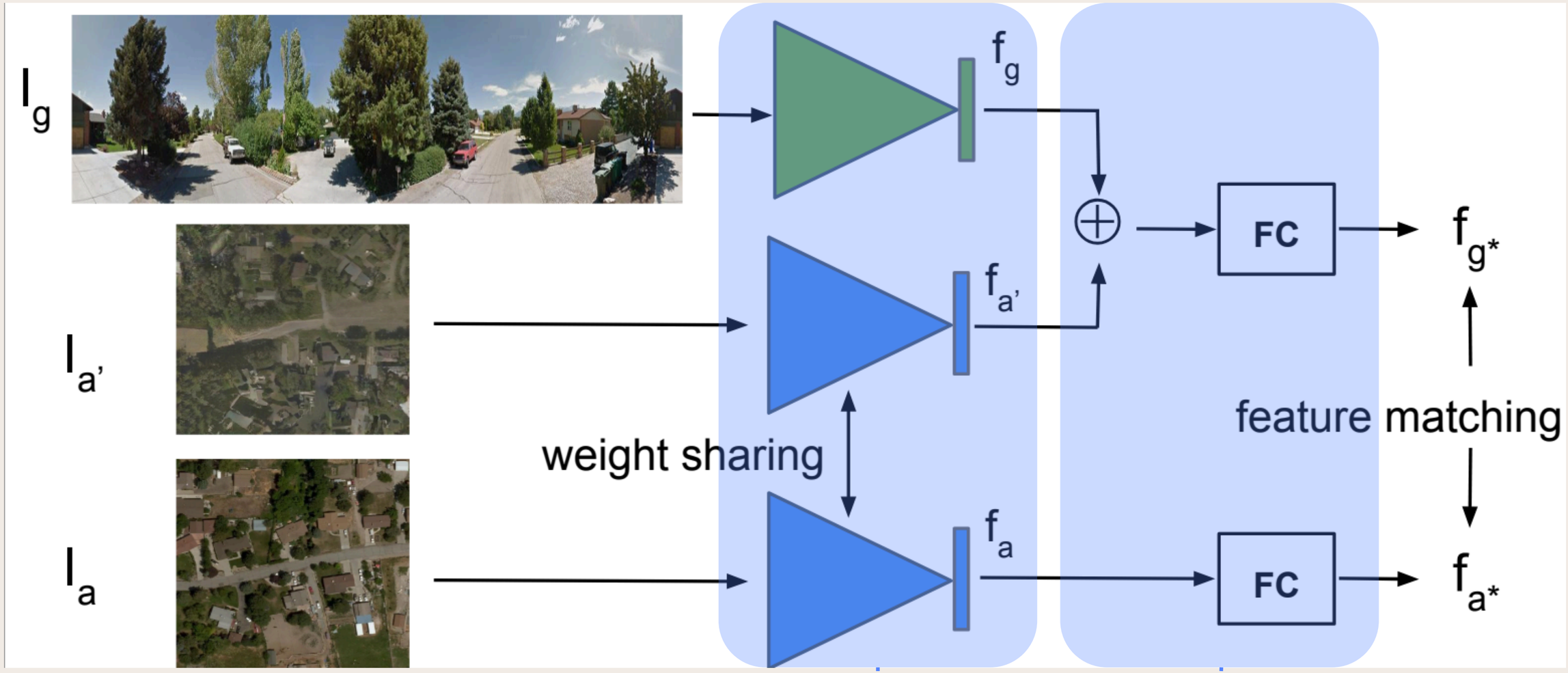
OUR APPROACH:

SEMANTIC SEGMENTATION

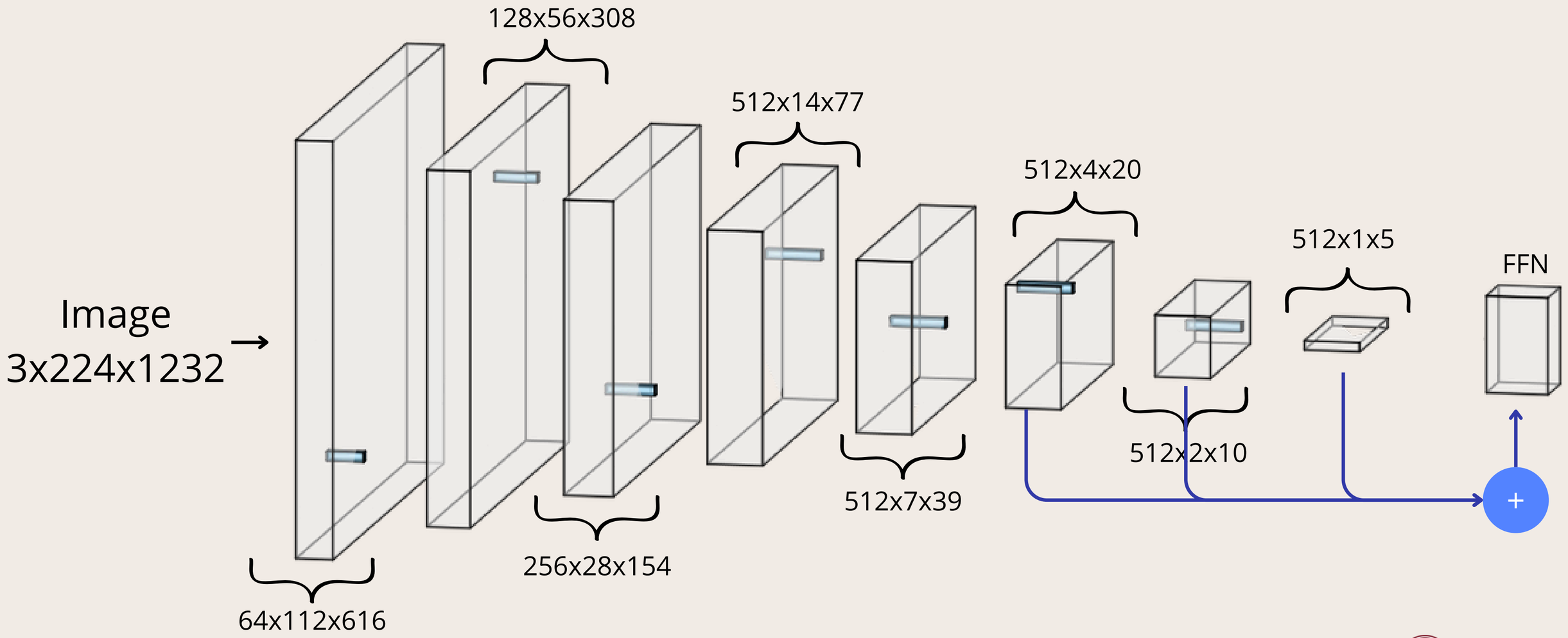
BUT WHY?

SAPIENZA
UNIVERSITÀ DI ROMA

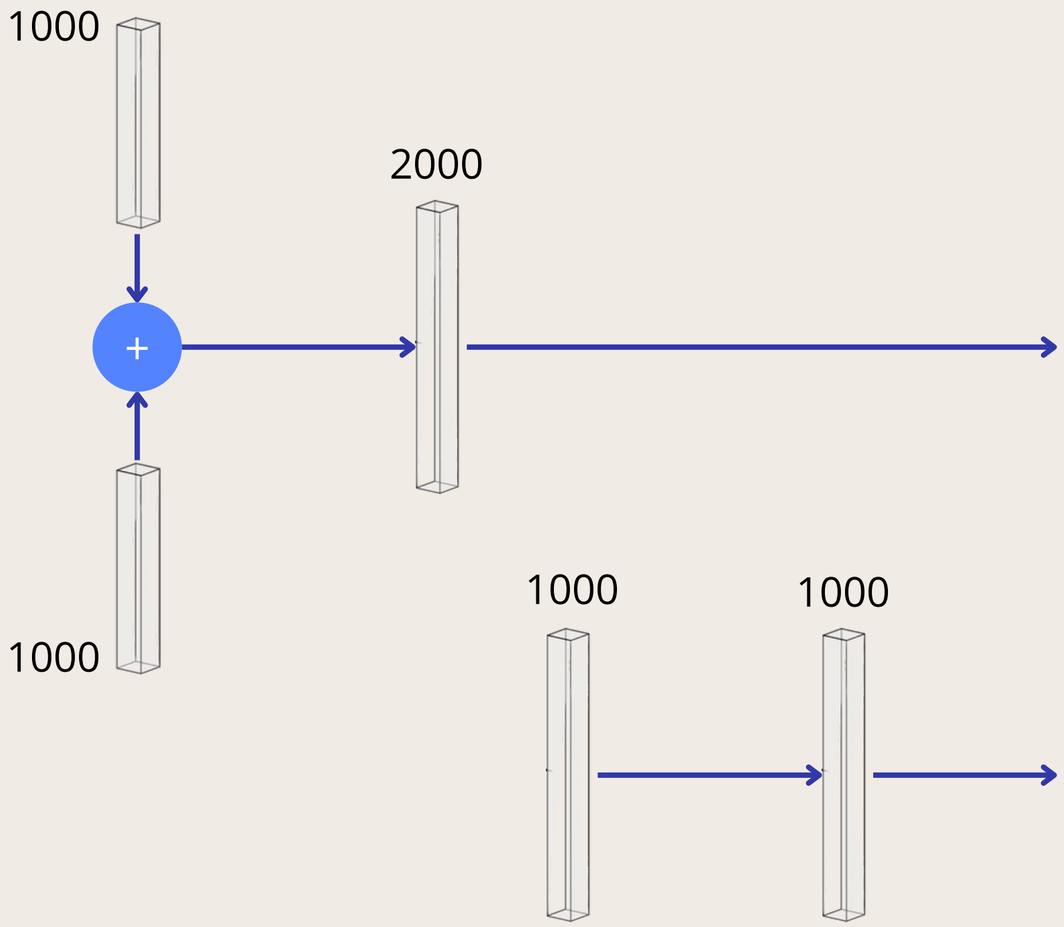
FULL NETWORK (Regmi et al. 2019)



JOINT FEATURE LEARNING NETWORK



FEATURE FUSION NETWORK



TRAINING AND EVALUATION

TRAINING PARAMETERS

DATASET	CVUSA
EPOCHS	30
BATCH S.	8
L.R	10e-5

TOP-K BASED EVALUATION

Method	Top-1	Top-10	Top-1%
Two-stream baseline ($I_{a'}$, I_a)	10.23%	35.10%	72.58%
Two-stream baseline (I_g, I_a)	18.45%	48.98%	82.94%
Joint Feat. Learning ($I_{a'}$, I_a)	14.31%	48.75%	86.47%
Joint Feat. Learning (I_g, I_a)	29.75%	66.34%	92.09%
Feature Fusion	**48.75%**	**81.27%**	**95.98%**

SAPIENZA
UNIVERSITÀ DI ROMA

THANK YOU!

MAFFONGELLI MATTIA, RAGO SALVATORE MICHELE