Scan2CapMMT

Dense Captioning for 3D Scenes with Transformers

Scan2CapMMT

I. Scan2Cap

II. Meshed-Memory Transformer

III. Scan2CapMMT

IV. Insights & First Results

V. Next Steps

Scan2CapMMT

I. Scan2Cap

II. Meshed-Memory Transformer

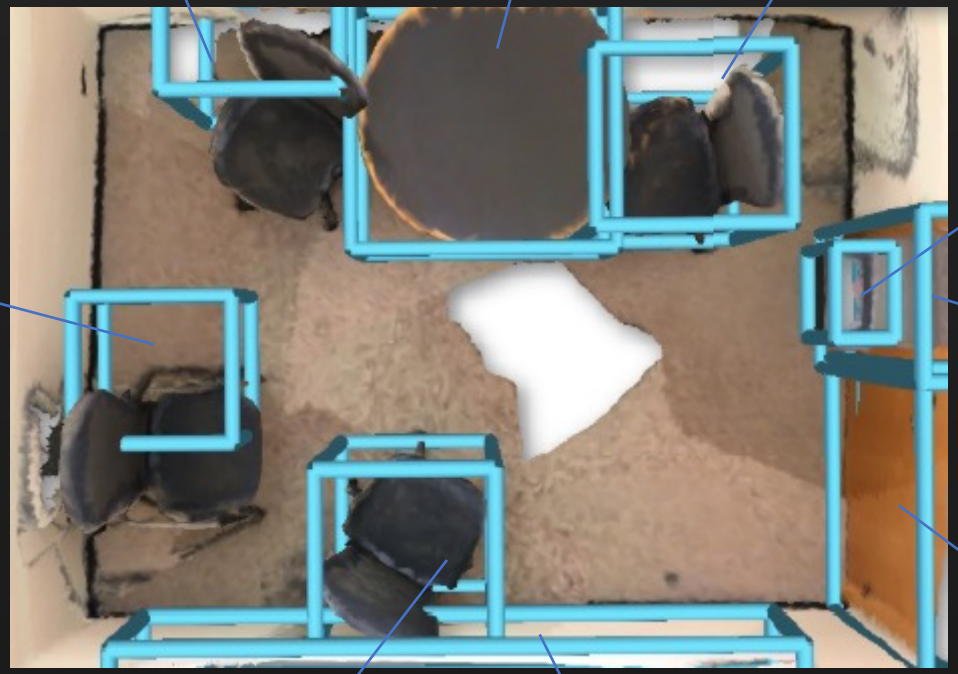
III. Scan2CapMMT

IV. Insights & First Results

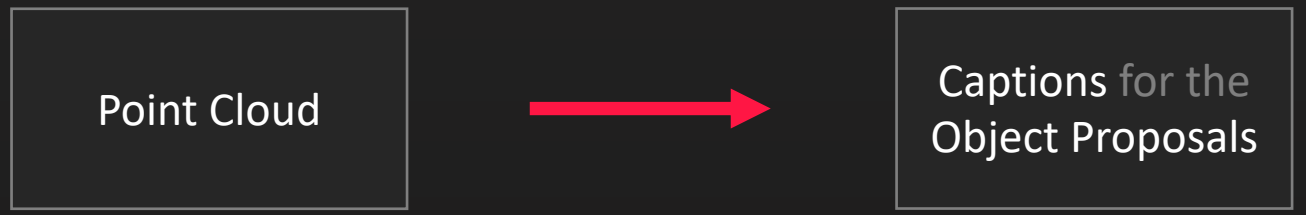
V. Next Steps

I. Scan2Cap: 3D Dense Captioning

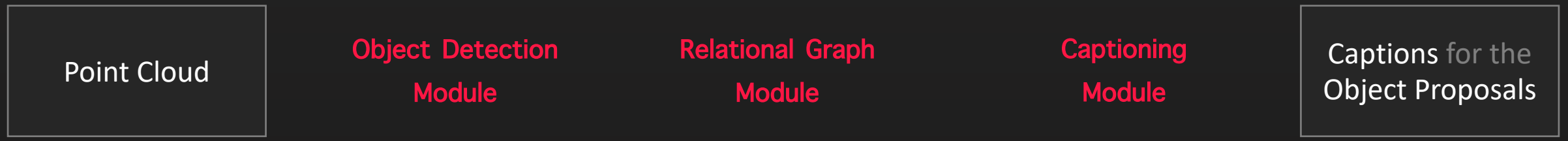
This is a black office chair. It is in the corner next to a black chair.



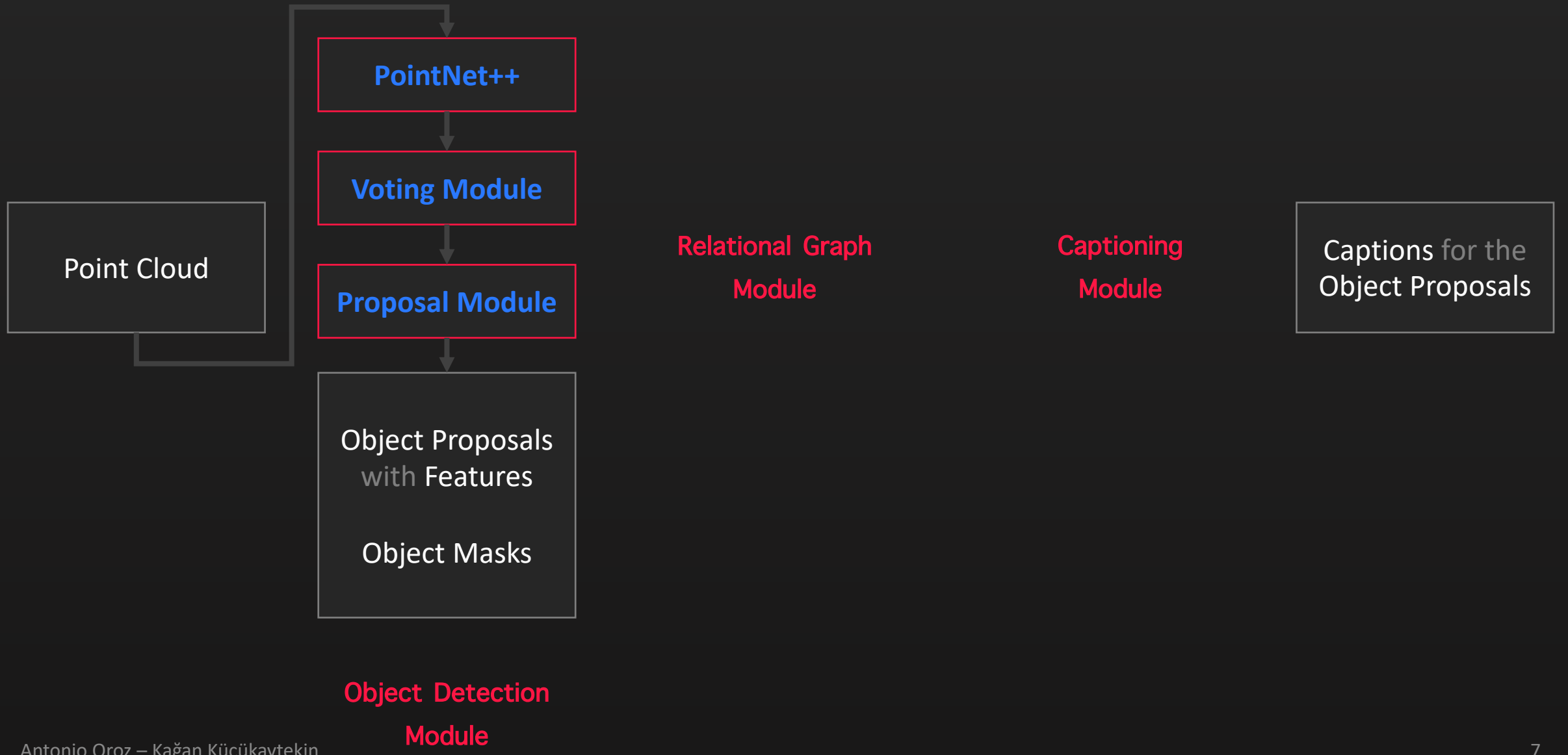
I. Scan2Cap



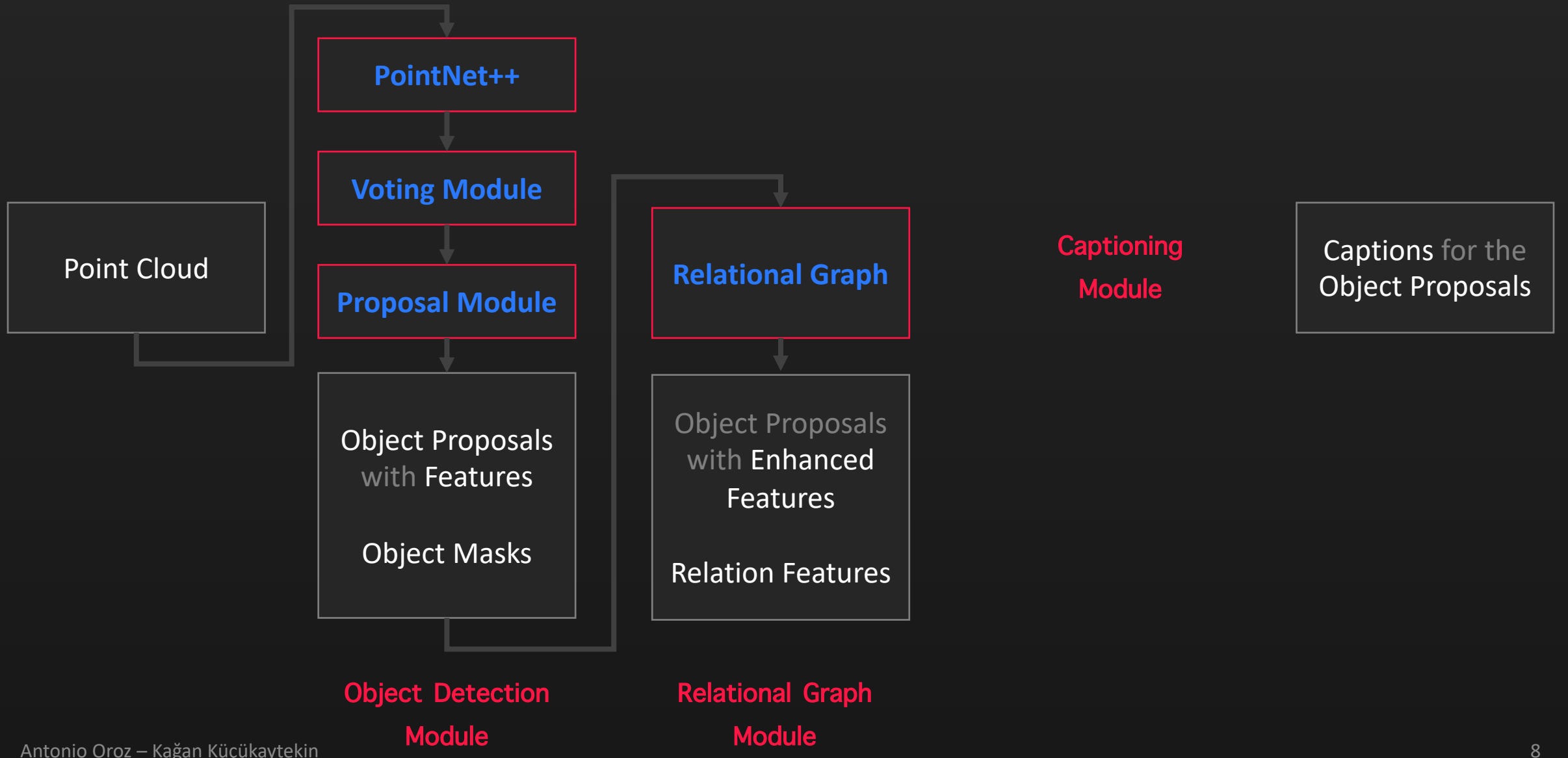
I. Scan2Cap: Architecture



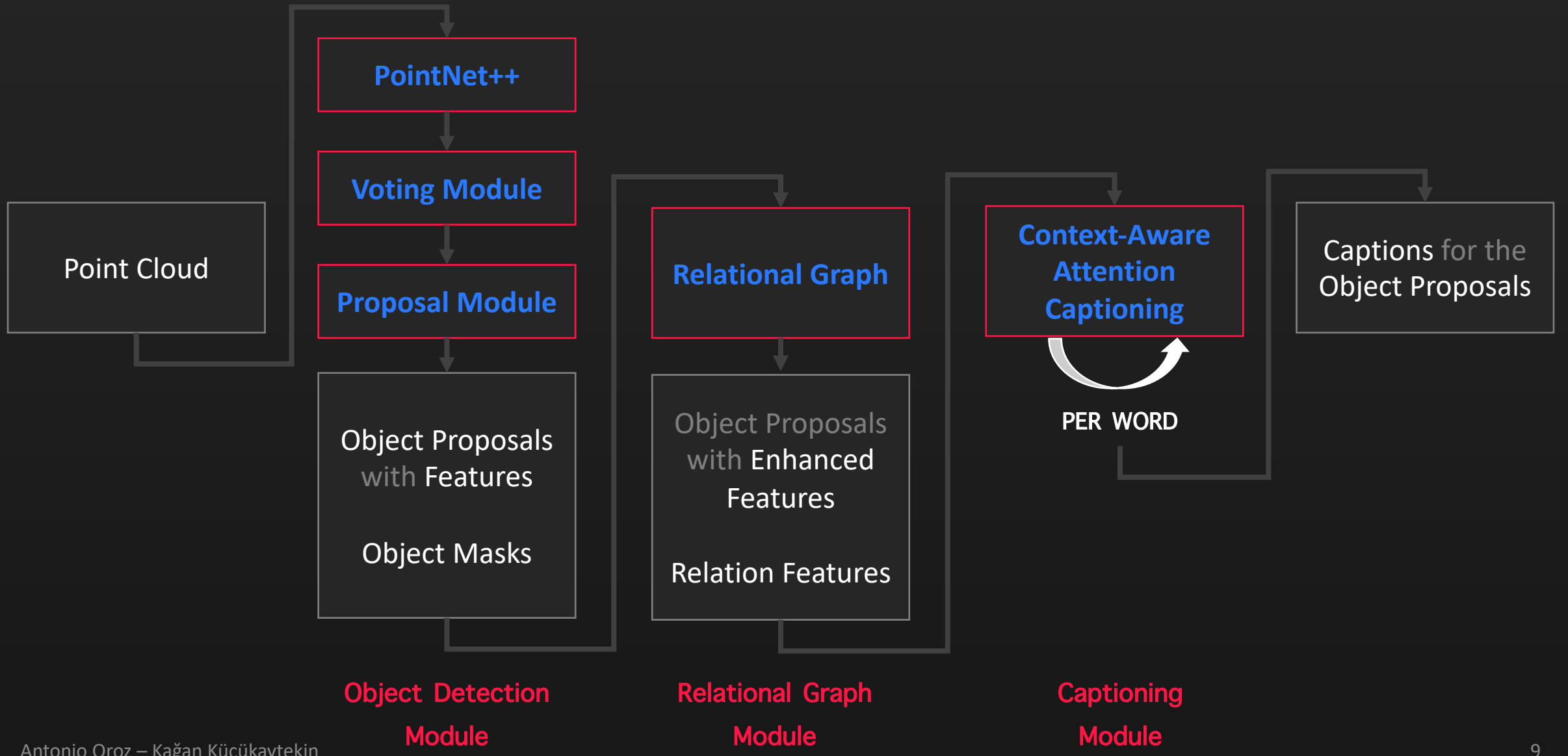
I. Scan2Cap: Architecture



I. Scan2Cap: Architecture



I. Scan2Cap: Architecture



Scan2CapMMT

I. Scan2Cap

II. Meshed-Memory Transformer

III. Scan2CapMMT

IV. Insights & First Results

V. Next Steps

Scan2CapMMT

I. Scan2Cap

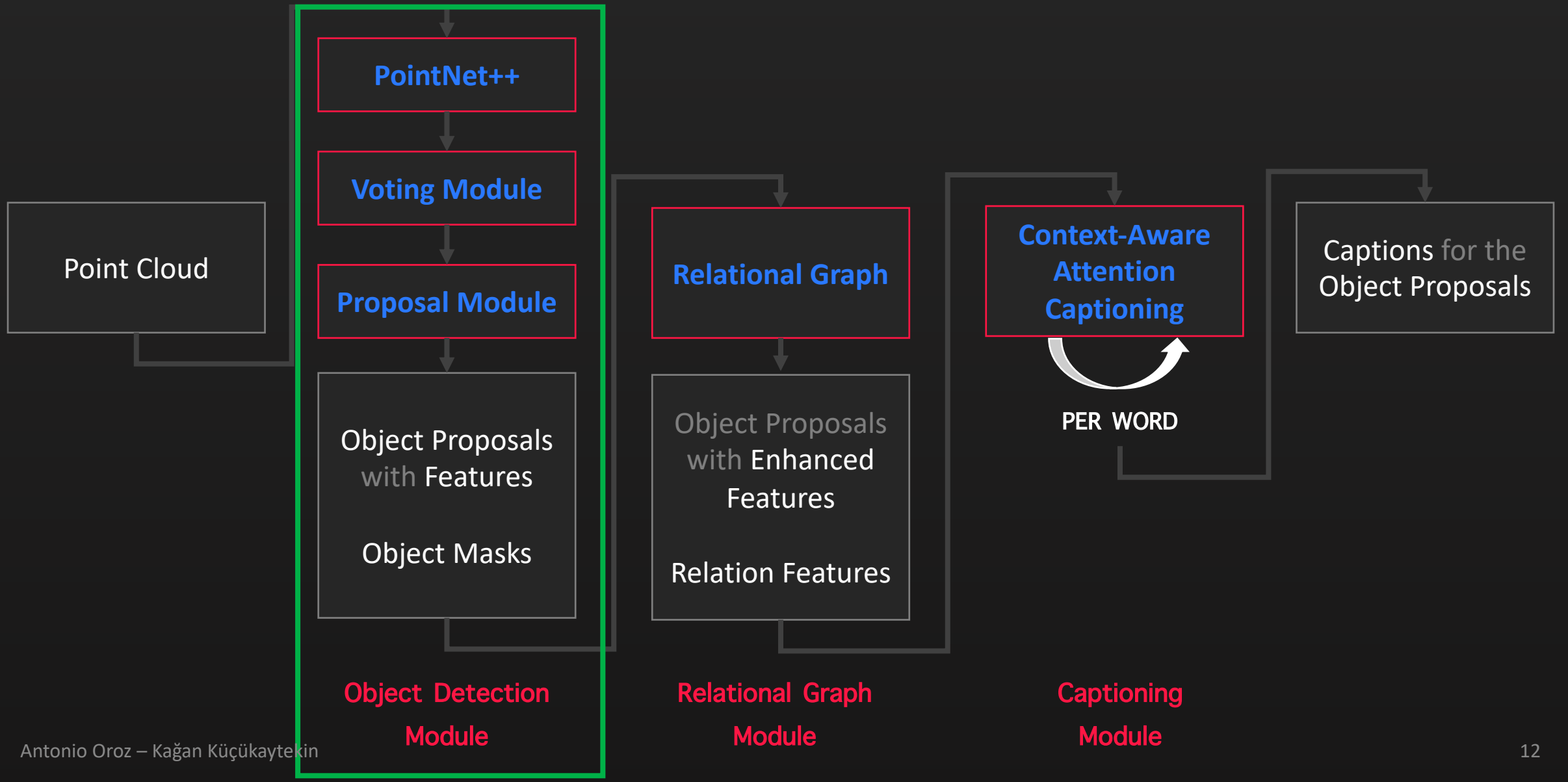
II. Meshed-Memory Transformer

III. Scan2CapMMT

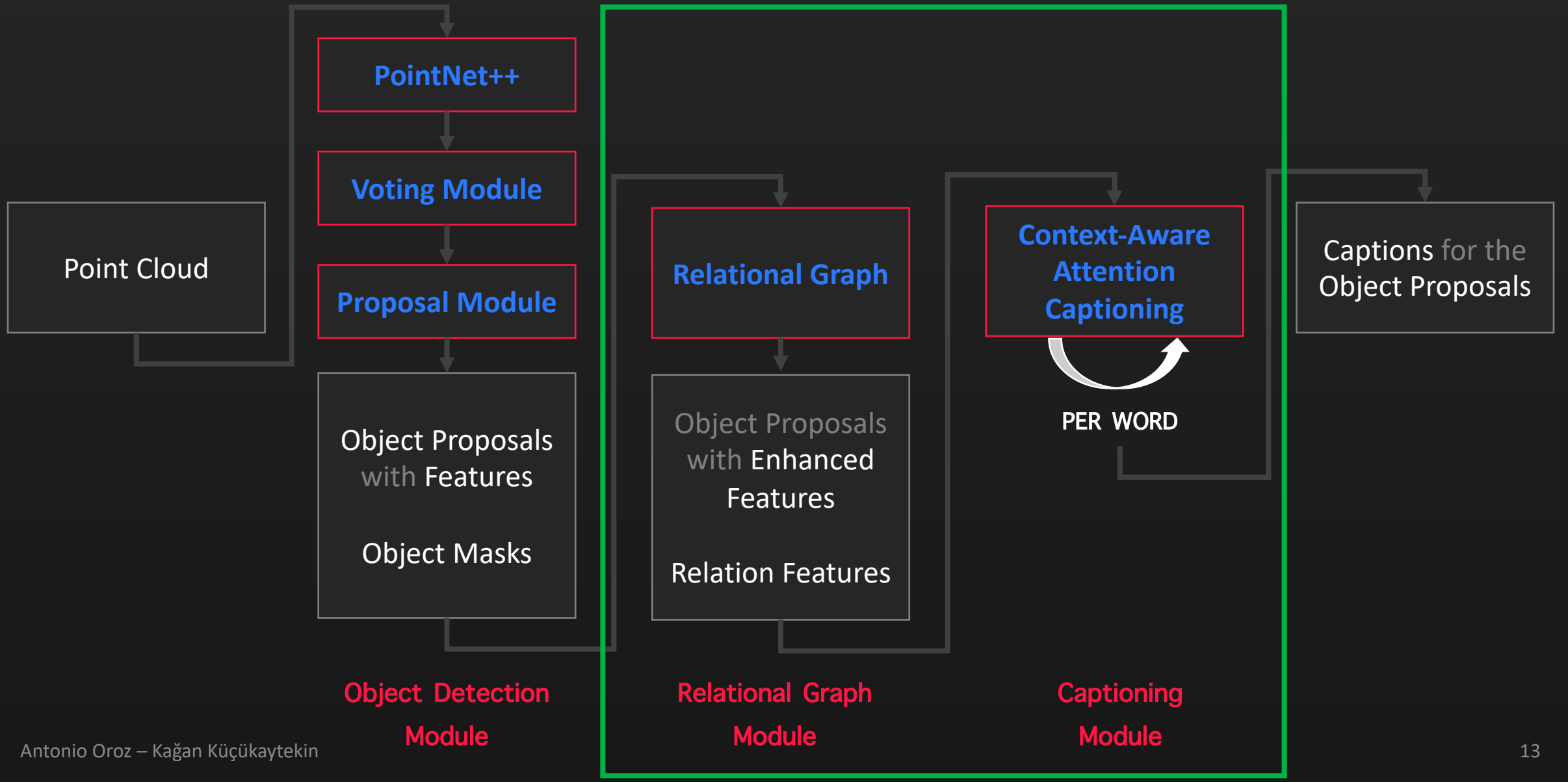
IV. Insights & First Results

V. Next Steps

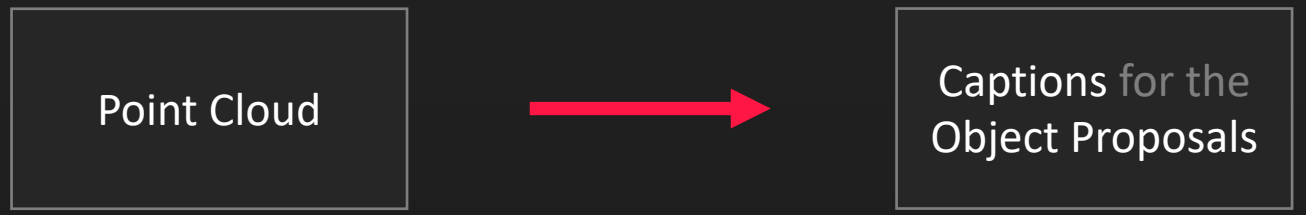
II. Scan2CapMMT: Motivation for MMT



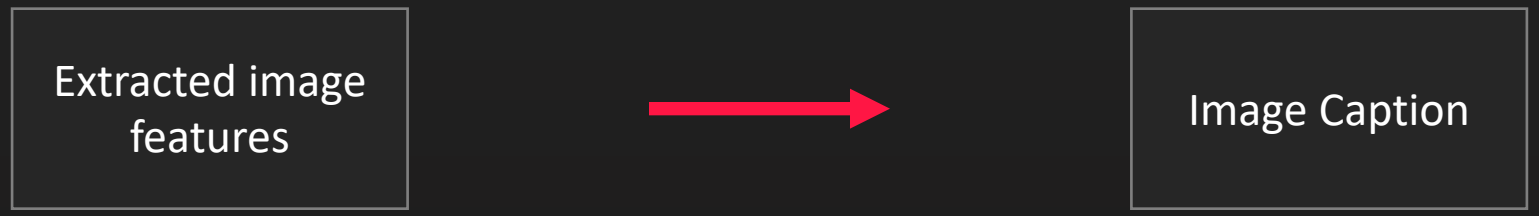
II. Scan2CapMMT: Motivation for MMT



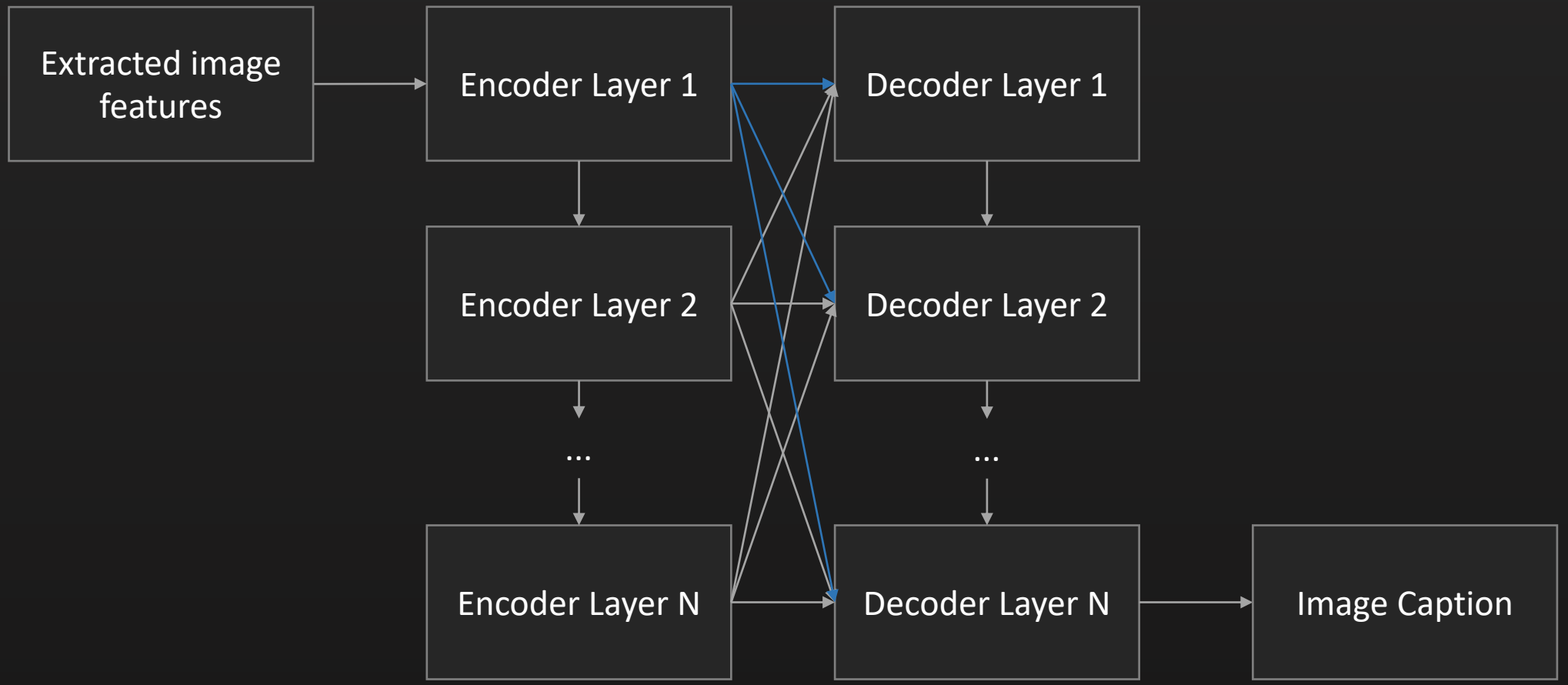
II. Scan2Cap



II. Meshed-Memory Transformer

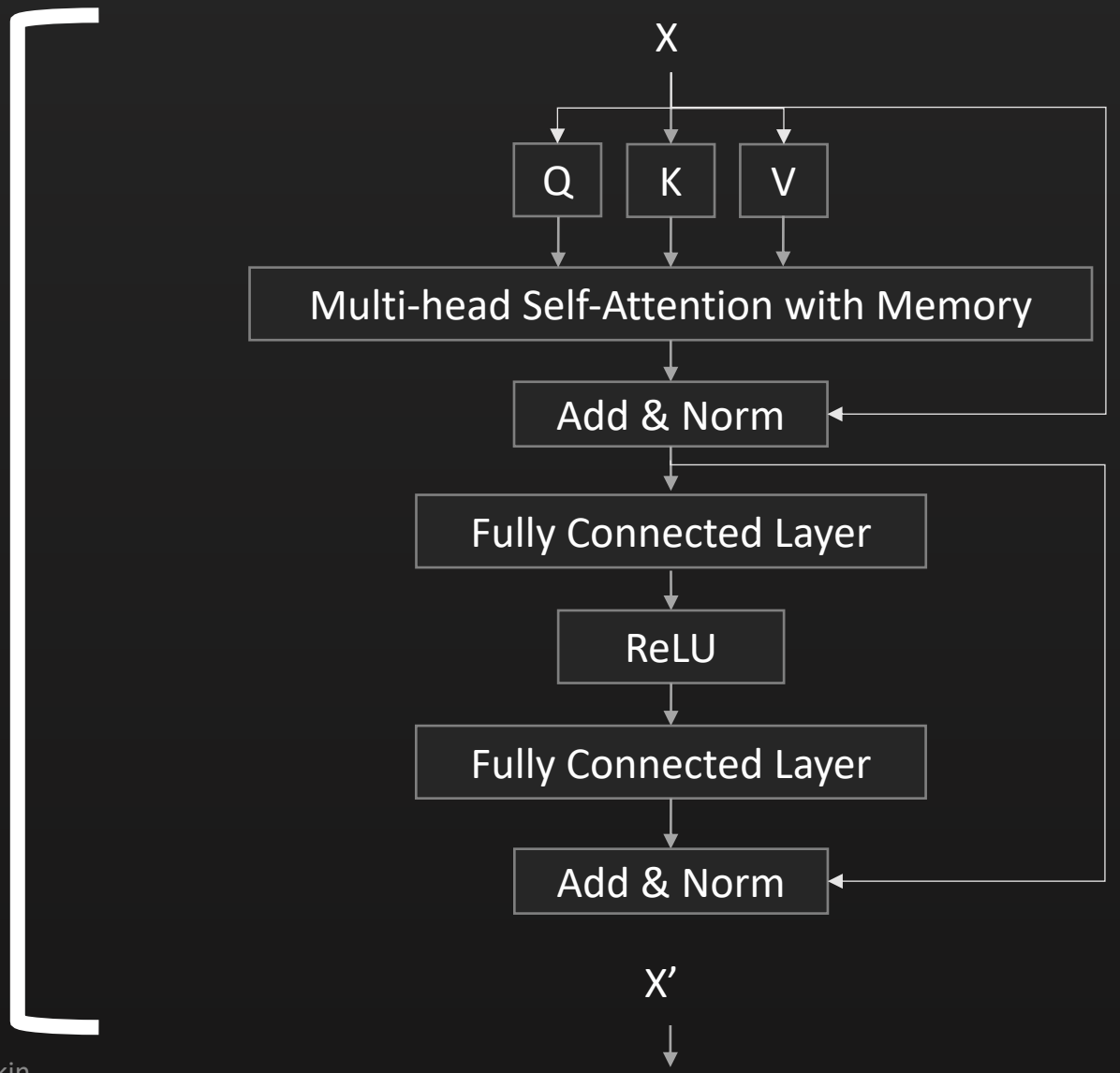


II. Meshed-Memory Transformer

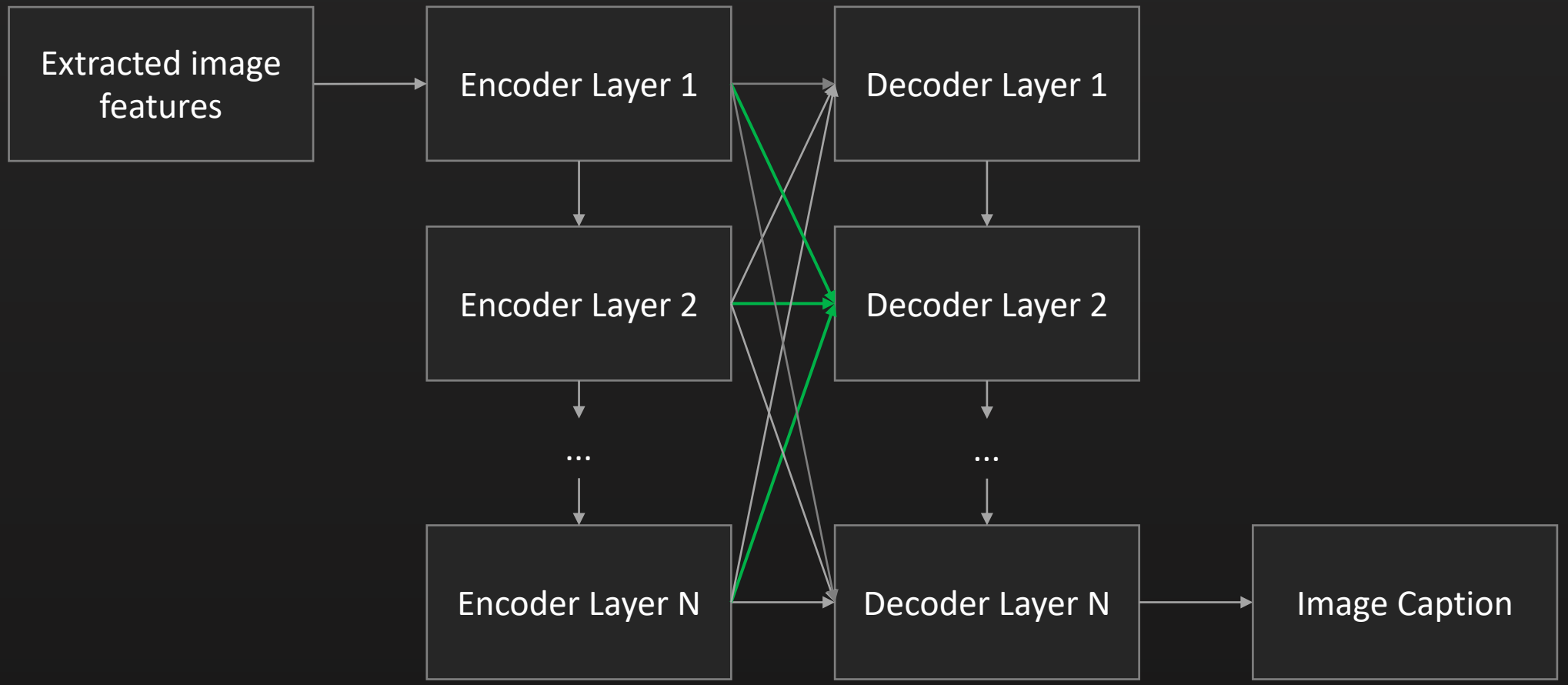


II. Meshed-Memory Transformer: Encoder

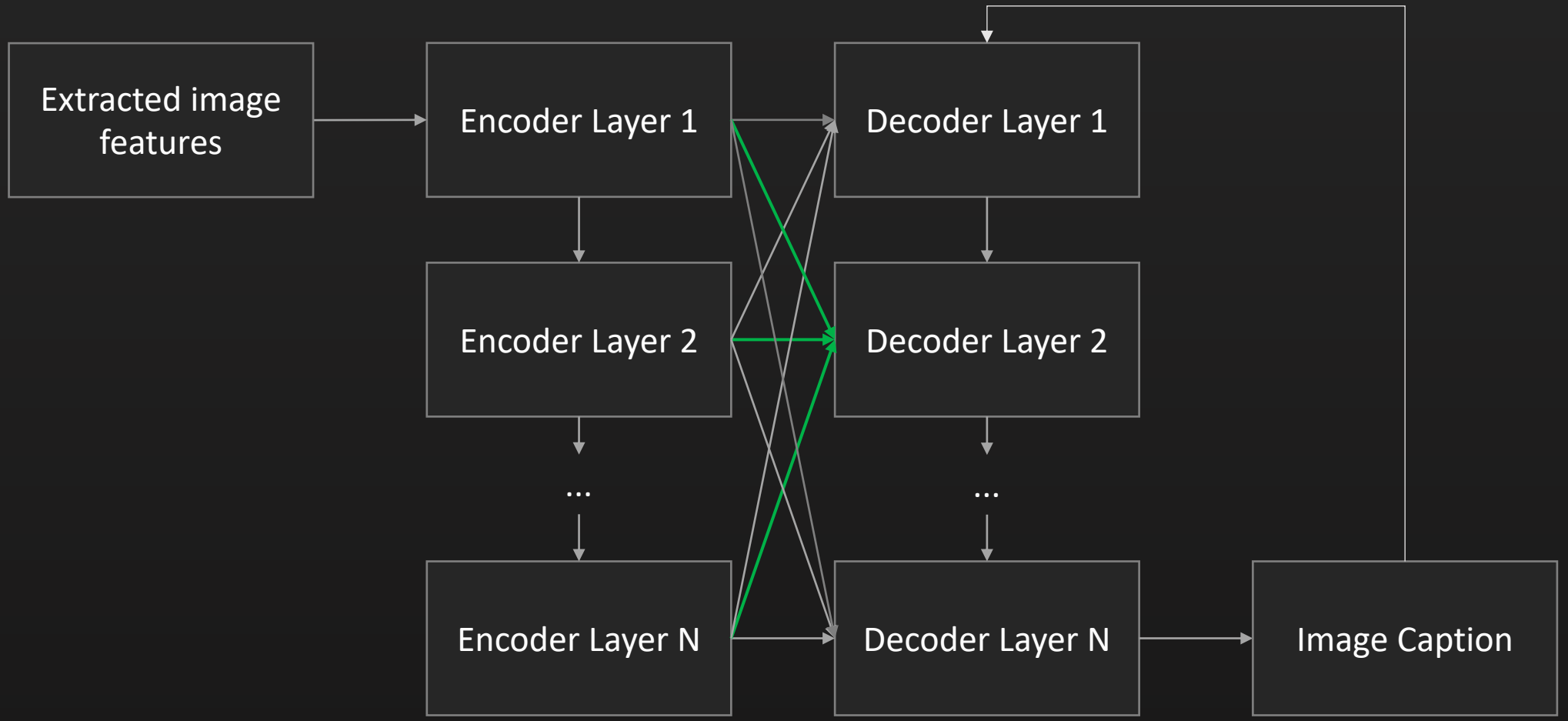
Encoder Layer 1



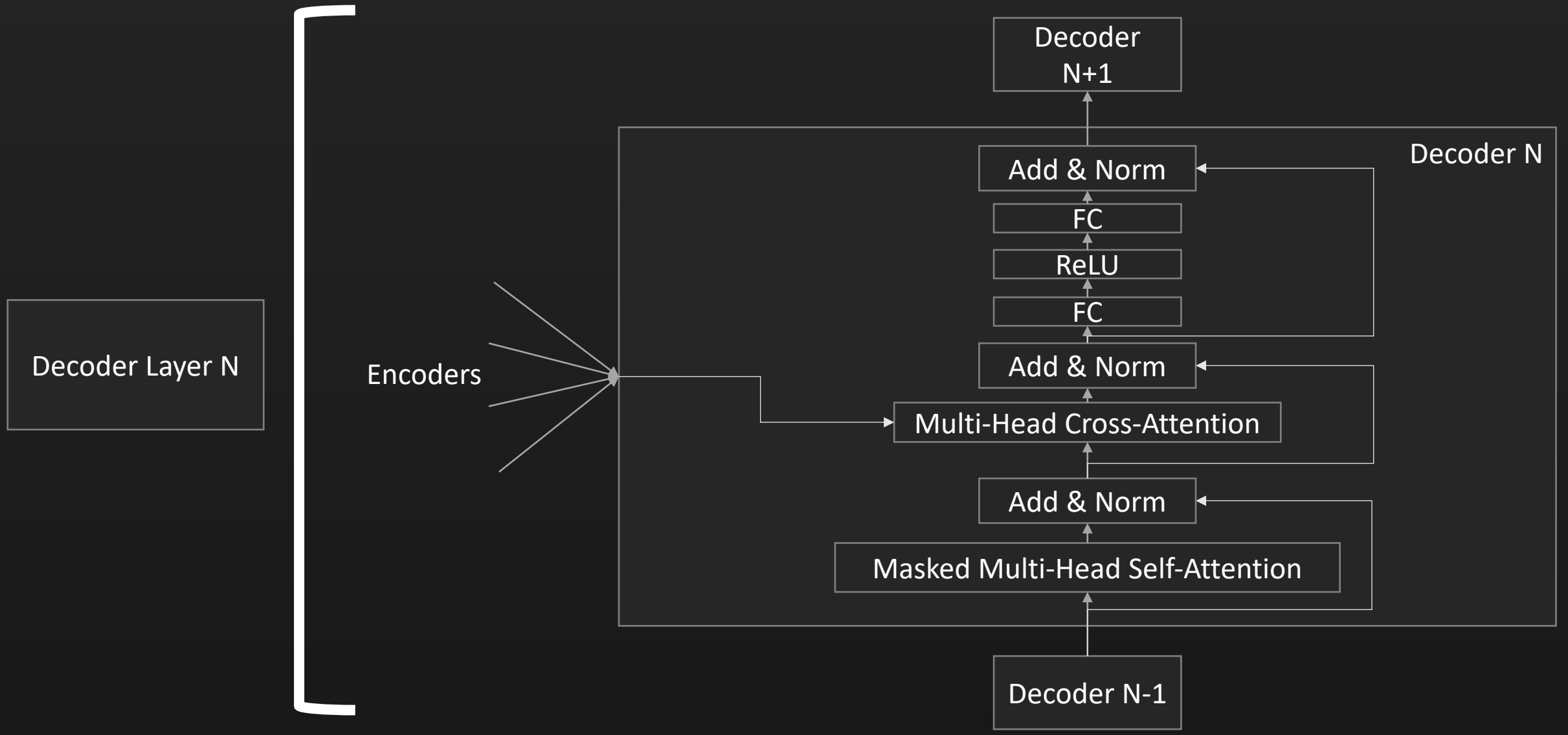
II. Meshed-Memory Transformer



II. Meshed-Memory Transformer



II. Meshed-Memory Transformer: Decoder



Scan2CapMMT

I. Scan2Cap

II. Meshed-Memory Transformer

III. Scan2CapMMT

IV. Insights & First Results

V. Next Steps

Scan2CapMMT

I. Scan2Cap

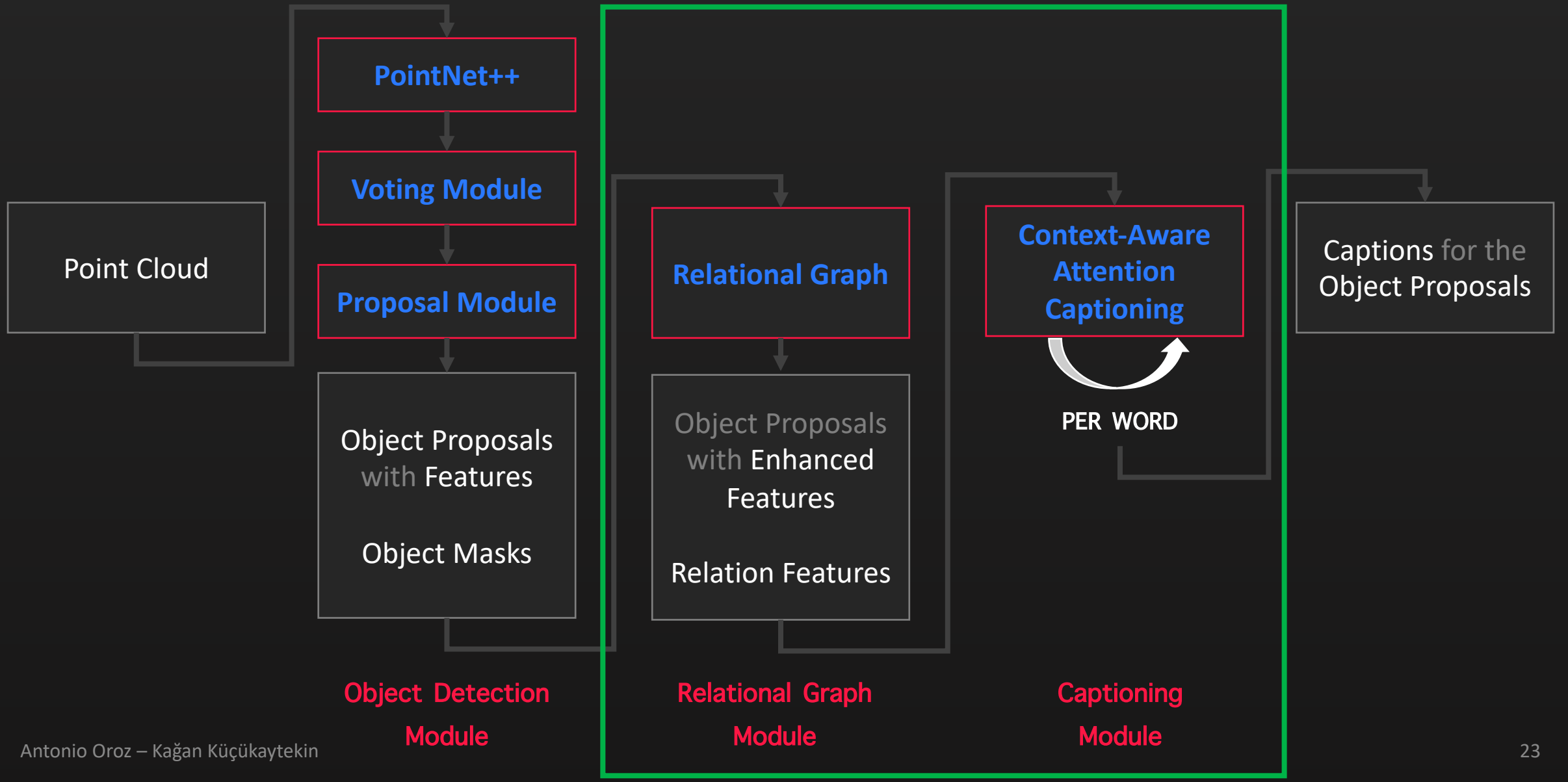
II. Meshed-Memory Transformer

III. Scan2CapMMT

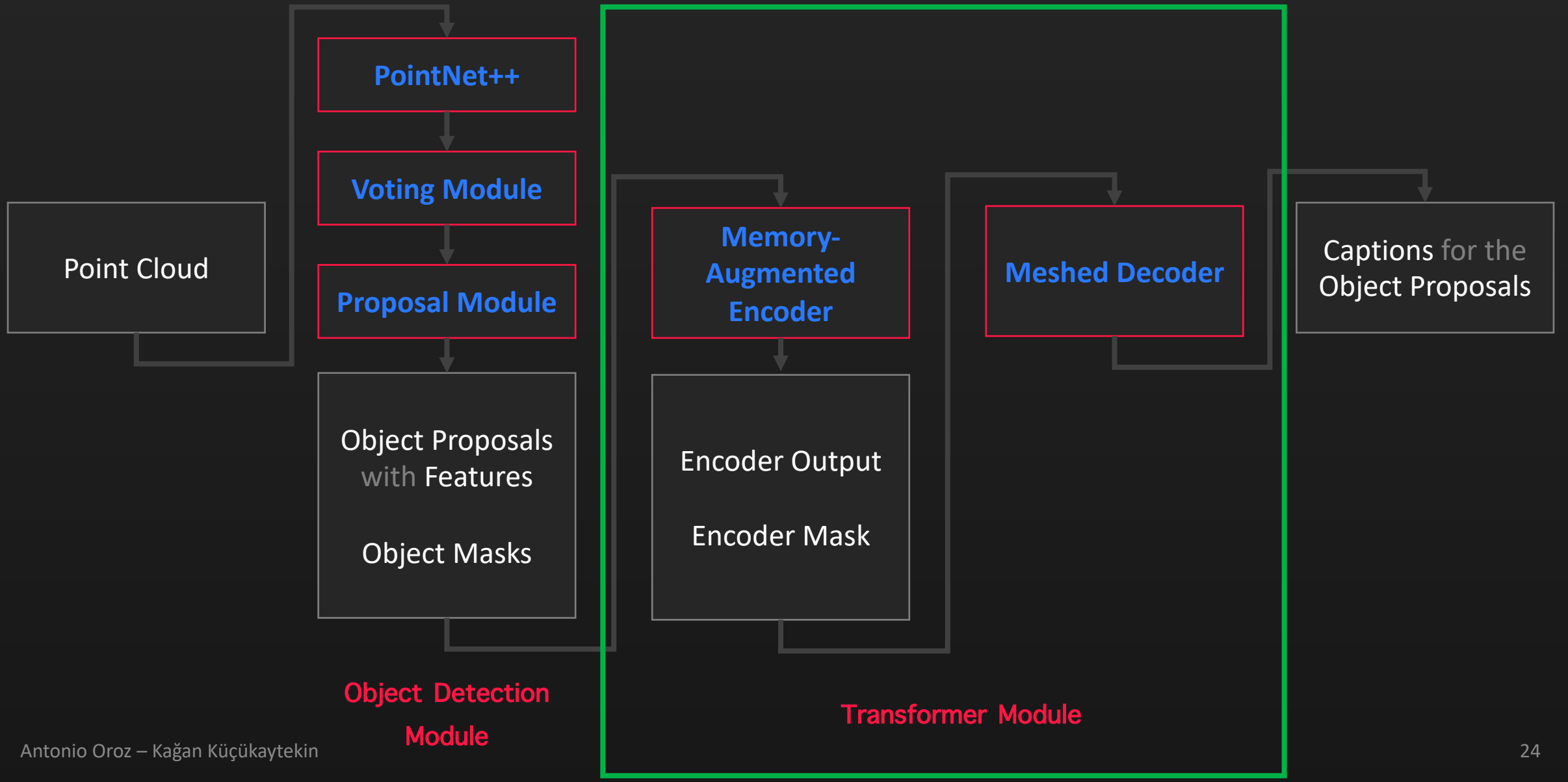
IV. Insights & First Results

V. Next Steps

III. Scan2CapMMT: Initial Architecture

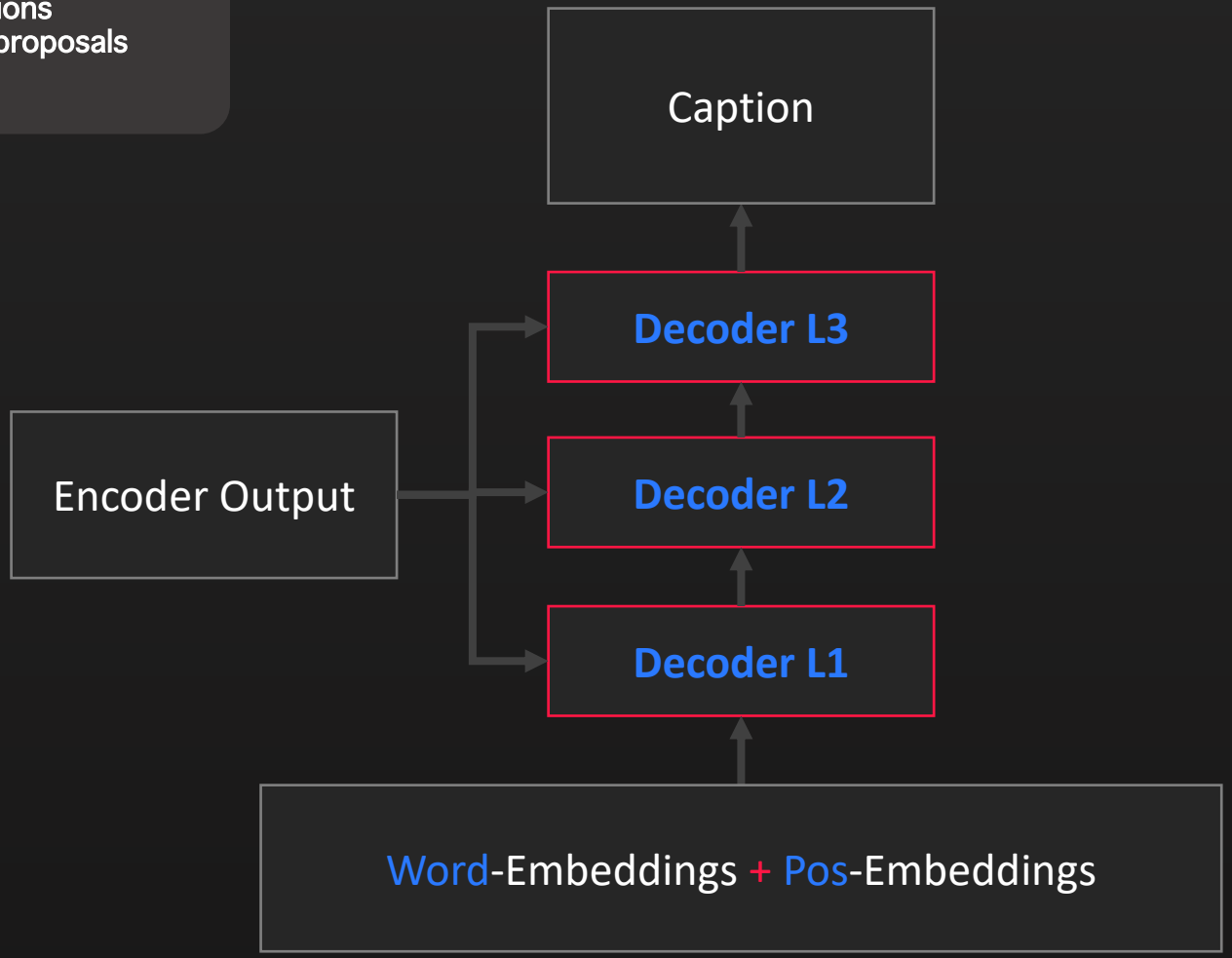


III. Scan2CapMMT: With MMT



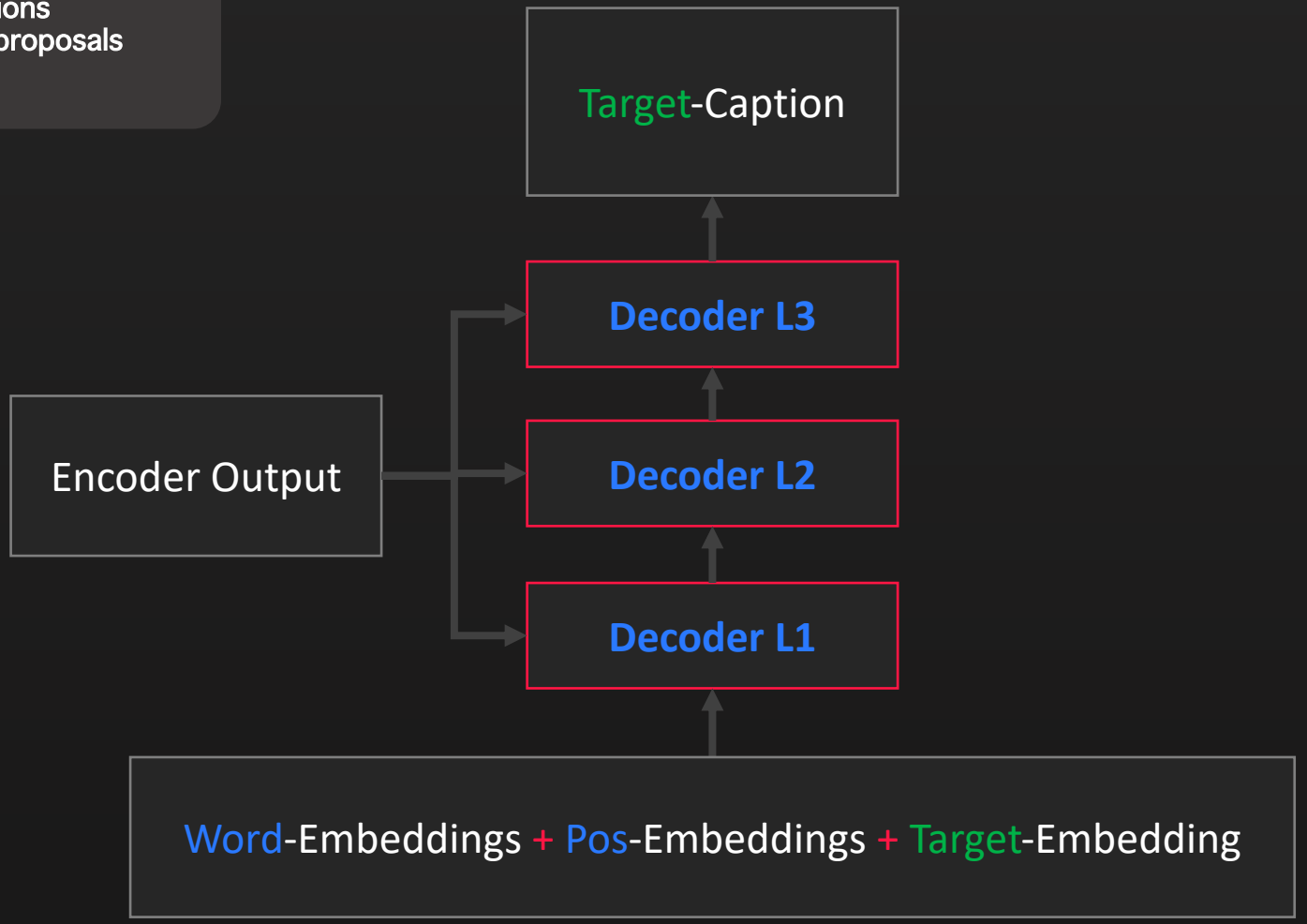
III. Scan2CapMMT: Challenges

Decoding Captions
for multiple object proposals



III. Scan2CapMMT: Challenges

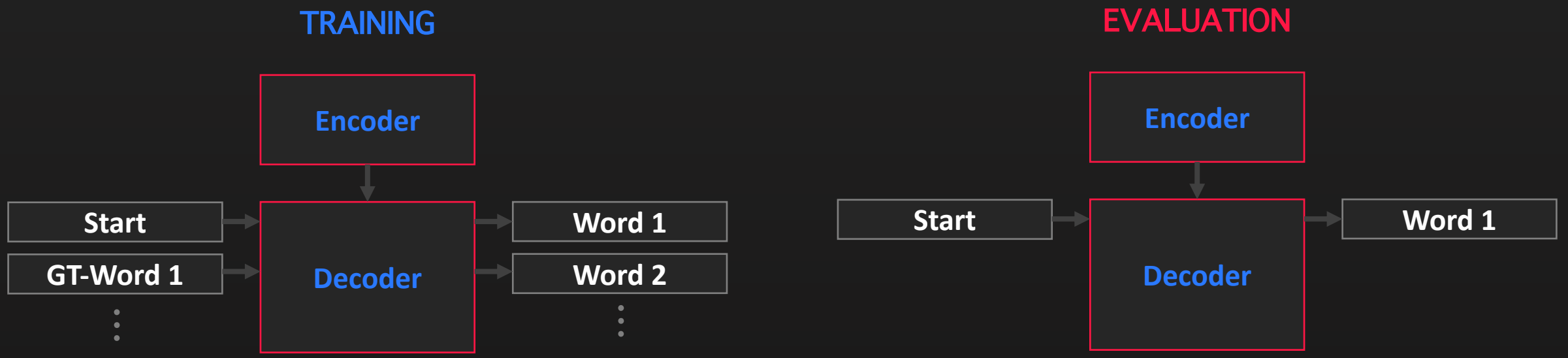
Decoding Captions
for multiple object proposals



III. Scan2CapMMT: Challenges

Decoding Captions
for multiple object proposals

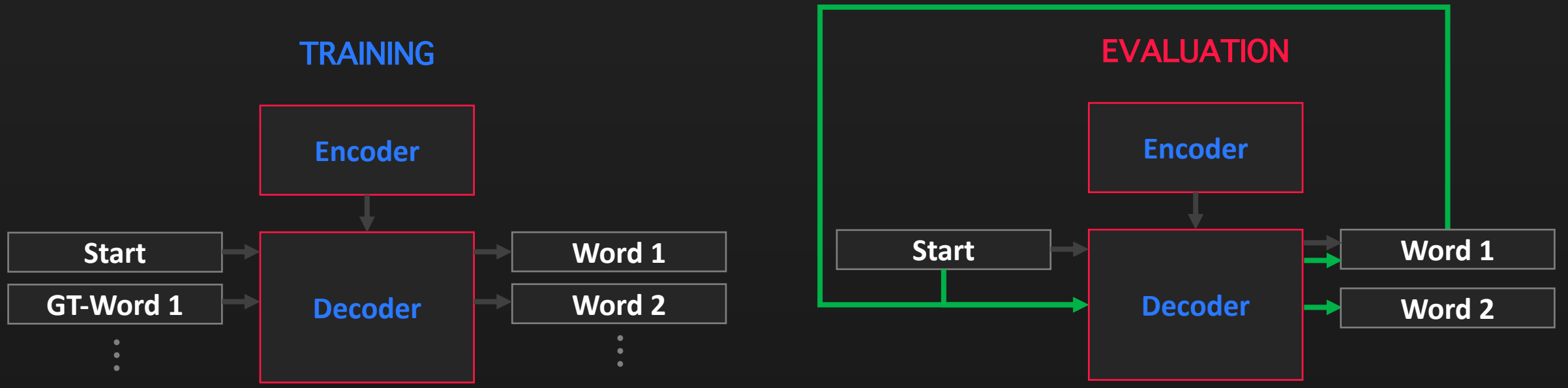
Caption-Generation
in Training and Evaluation



III. Scan2CapMMT: Challenges

Decoding Captions
for multiple object proposals

Caption-Generation
in Training and Evaluation



Scan2CapMMT

I. Scan2Cap

II. Meshed-Memory Transformer

III. Scan2CapMMT

IV. Insights & First Results

V. Next Steps

Scan2CapMMT

I. Scan2Cap

II. Meshed-Memory Transformer

III. Scan2CapMMT

IV. Insights & First Results

V. Next Steps

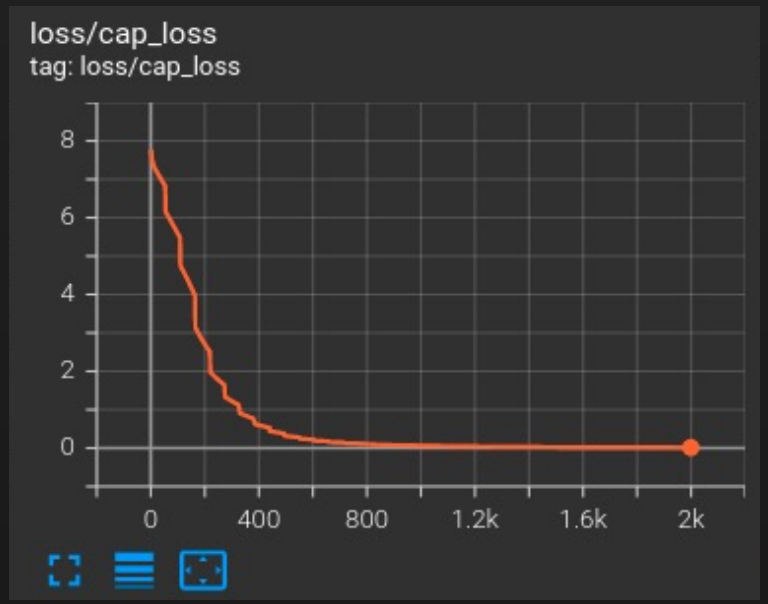
IV. Insights & First Results:

- Parameters: Scan2Cap MMT **7,830,308** vs **6,175,612** Scan2Cap
- Dropout: 0
- Weight Decay: 0
- Learning Rate: Changed from 1e-3 to 1e-4

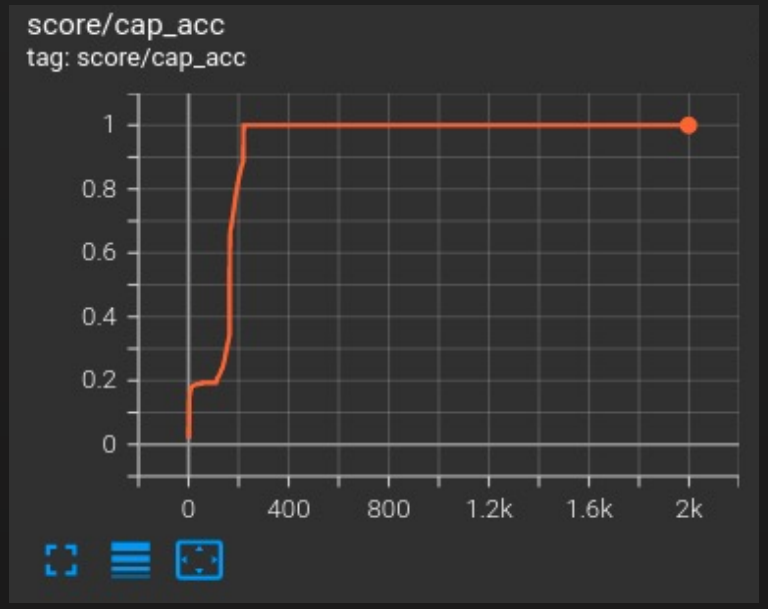
IV. Insights & First Results: Overfitting Results

1 SAMPLE 1 SCENE

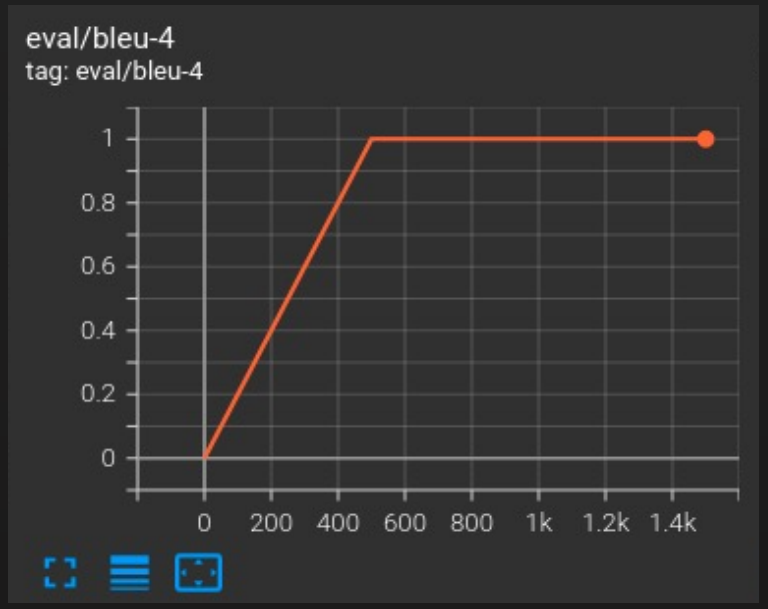
Caption Loss



Caption Accuracy



BLEU-4 Score

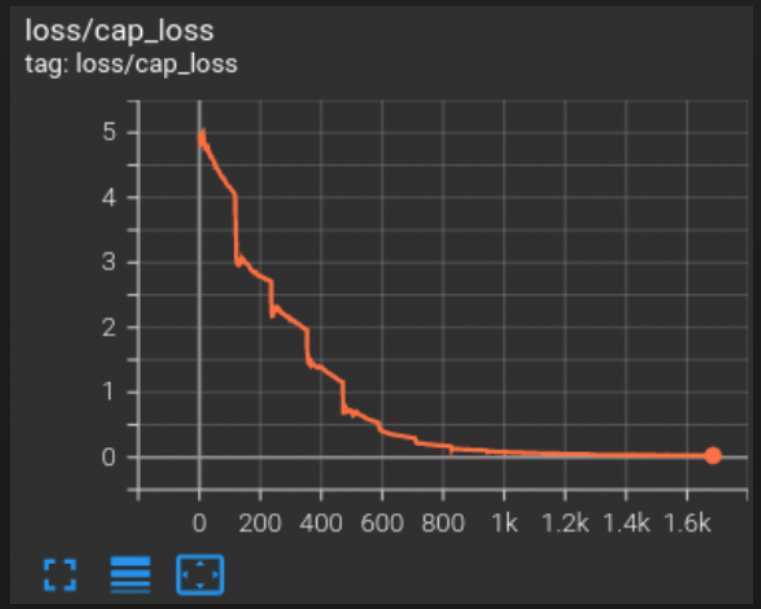


IV. Insights & First Results: Overfitting Results

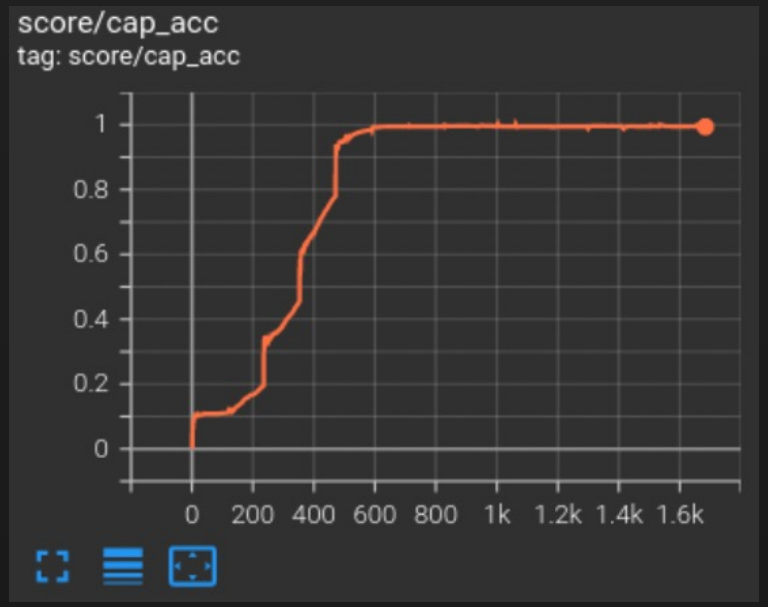
1 SAMPLE 1 SCENE

N SAMPLES 1 SCENE

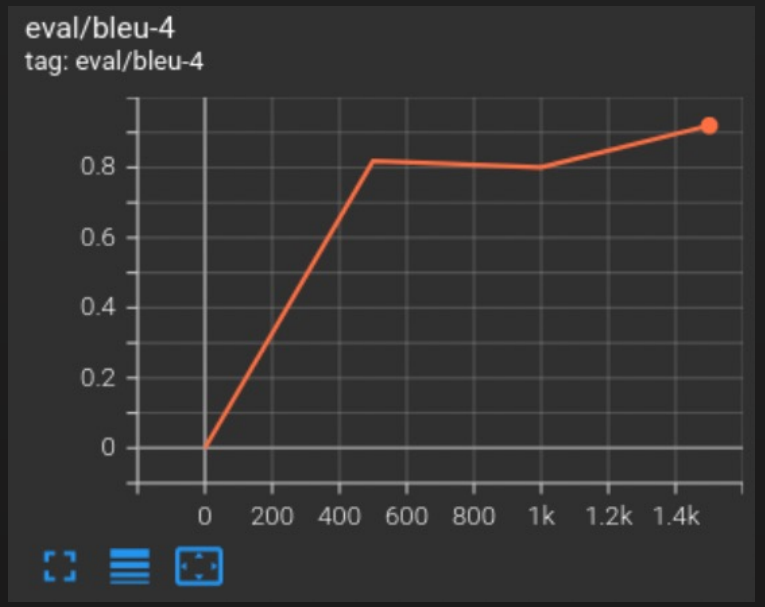
Caption Loss



Caption Accuracy



BLEU-4 Score



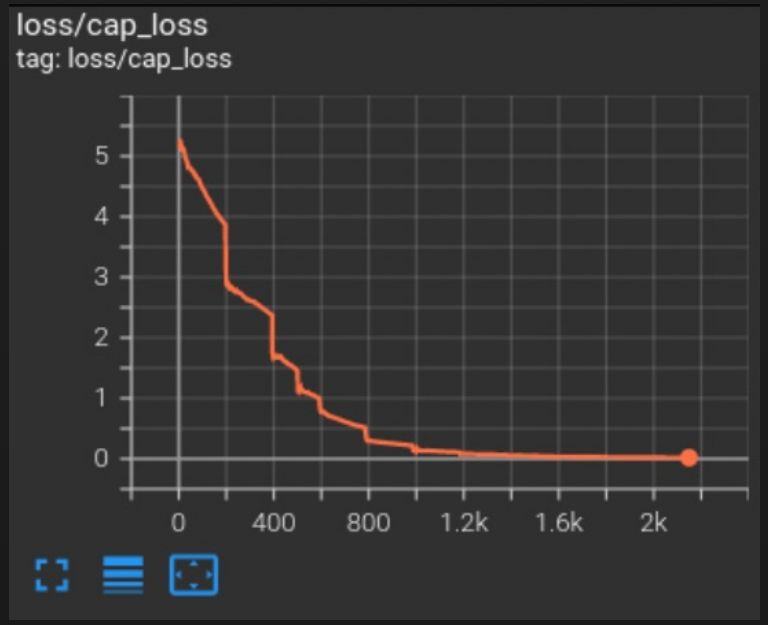
IV. Insights & First Results: Overfitting Results

1 SAMPLE 1 SCENE

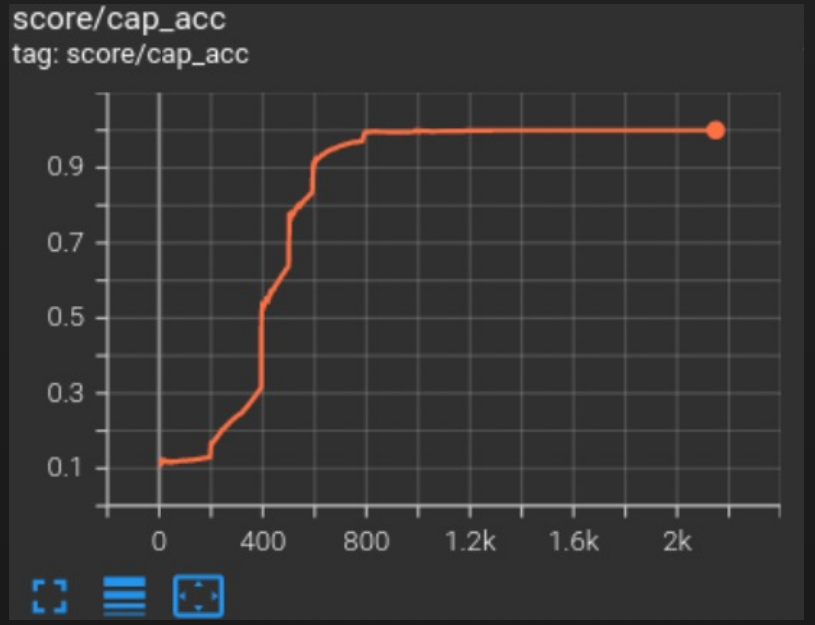
N SAMPLES 1 SCENE

N SAMPLES M SCENES

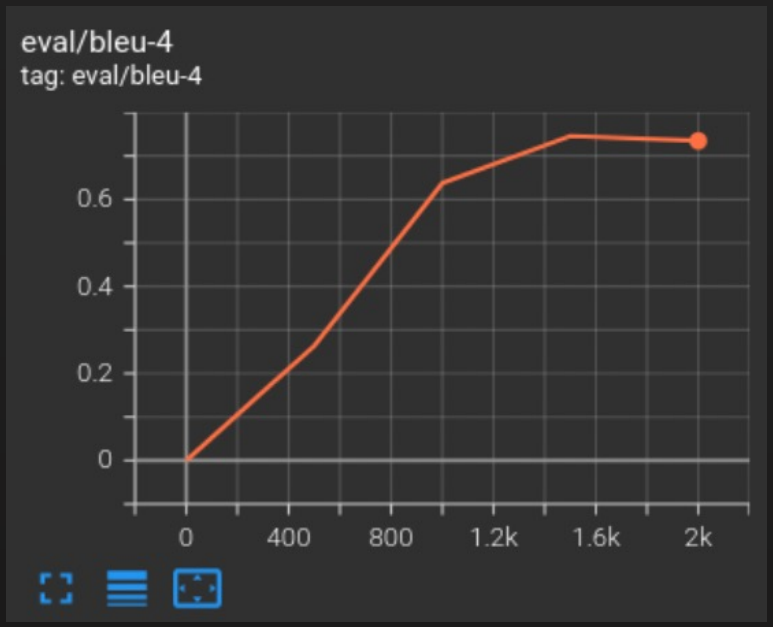
Caption Loss



Caption Accuracy



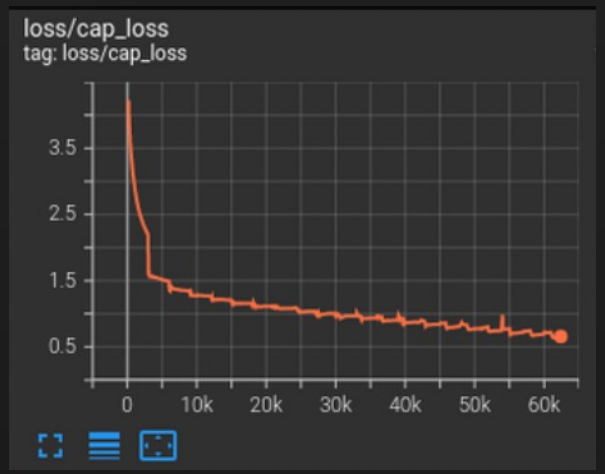
BLEU-4 Score



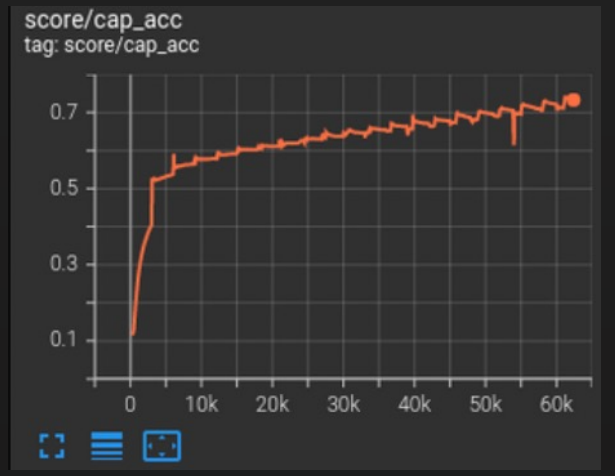
IV. Insights & First Results: Training on the whole Dataset

Losses & Accuracies

Caption Loss



Caption Accuracy



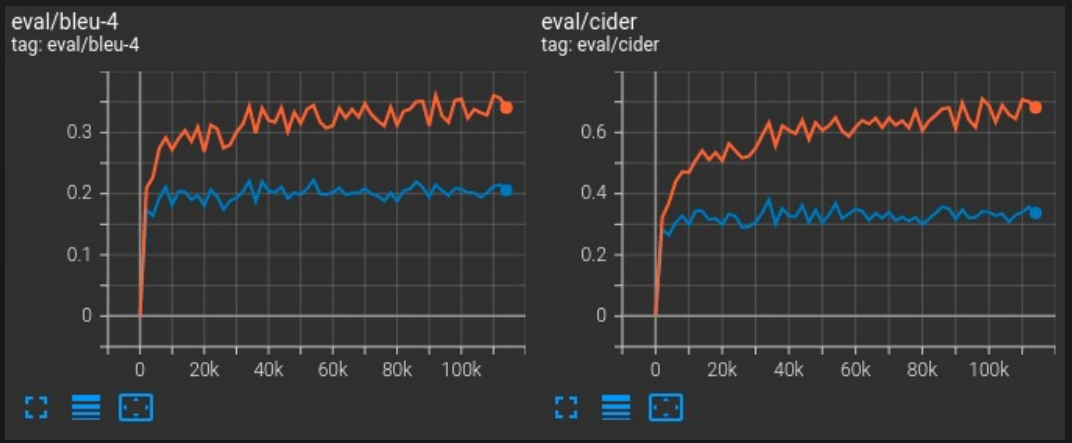
IV. Insights & First Results: Training on the whole Dataset

Losses & Accuracies

Evaluation

Scan2CapMMT

Scan2Cap



Scan2CapMMT

I. Scan2Cap

II. Meshed-Memory Transformer

III. Scan2CapMMT

IV. Insights & First Results

V. Next Steps

Scan2CapMMT

I. Scan2Cap

II. Meshed-Memory Transformer

III. Scan2CapMMT

IV. Insights & First Results

V. Next Steps

V. Next Steps

BEAM SEARCH

Instead of generating one sentence for an object proposal,
generate multiple sentences in parallel and choose the final sentence
with log propobabilities.

V. Next Steps

BEAM SEARCH

REINFOCEMENT LEARNING

After pretraining on the Cross-Entropy loss,
use Reinforcement Learning with CIDEr-D as a reward
to train the model.

V. Next Steps

BEAM SEARCH

REINFOCEMENT LEARNING

HYPERPARAMETER TUNING

- Internal Dimensions of MMT
- Number of Proposals
- # Decoder-/Encoder-Layers
- Schedules
- Learning Rate
- Weight Decay

...

V. Next Steps

BEAM SEARCH | REINFOCEMENT LEARNING | HYPERPARAMETER TUNING | GROUP-FREE TRANSFORMER

Replace the current detection module
with the Group-Free 3D Object Detection via Transformers module
proposed by Liu et al.

V. Next Steps

BEAM SEARCH | REINFOCEMENT LEARNING | HYPERPARAMETER TUNING | GROUP-FREE TRANSFORMER | **AoA**

MMT currently uses Dot-Product Attention
which we could replace with Attention on Attention

Scan2CapMMT

I. Scan2Cap

II. Meshed-Memory Transformer

III. Scan2Cap with MMT

IV. Insights & First Results

V. Next Steps

Scan2CapMMT

THANK YOU FOR YOUR ATTENTION :D