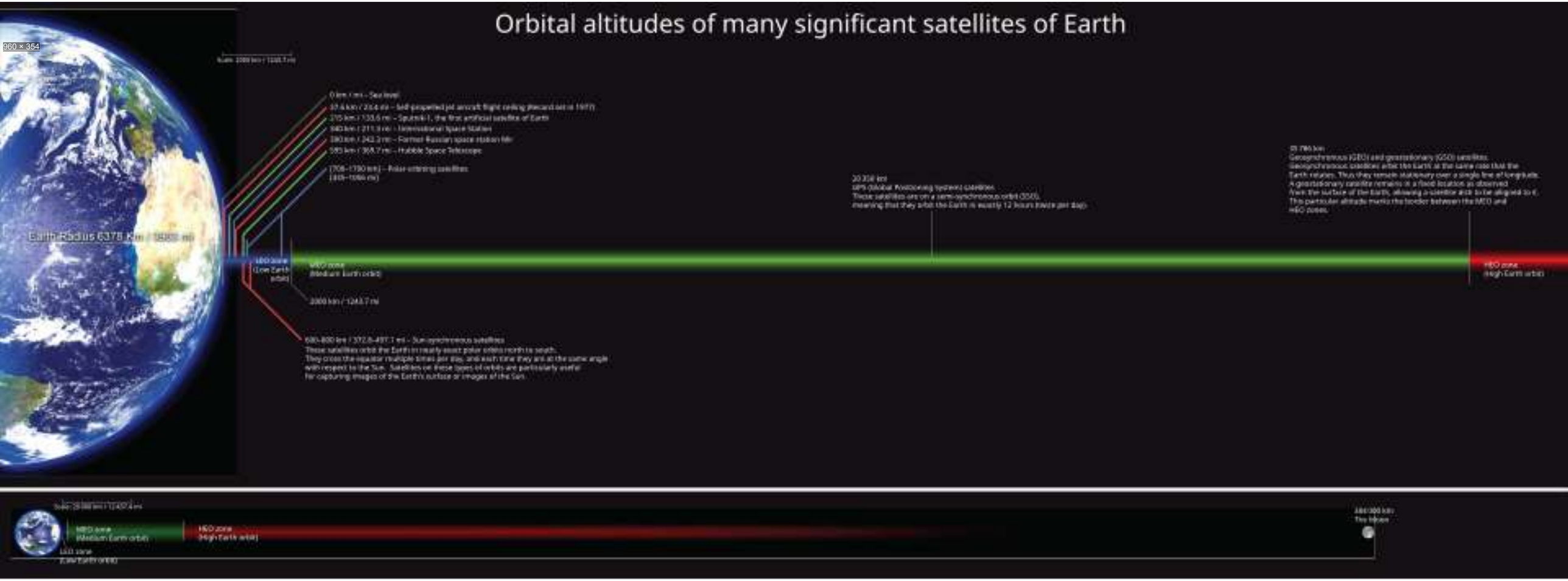

澳門大學
UNIVERSIDADE DE MACAU
UNIVERSITY OF MACAU

# Multimedia UAVs: Capturing the World from a New Perspective

**Zhedong Zheng**
University of Macau

# Aerial Space: From altitude view




# Where is Starlink?

# Aerial Space: From altitude view


# Where is Starlink?

# Comparison of Satellite, HAPS, and UAV Platforms

| Parameter | Satellite | HAPS | UAV |
|---|---|---|---|
| **Altitude** | 500–36,000 km | 20–50 km | < 10 km |
| **Spatial Resolution** | m to sub-m | dm to cm | cm-level |
| **Temporal Coverage** | Hours–days revisit | Persistent (regional) | Real-time (local) |
| **Coverage Area** | Global/regional | Regional | Localized |
| **Endurance** | Years | Months | Hours |
| **Maneuverability** | Minimal | Limited | High |
| **Data Latency** | High | Medium | Low |
| **Payload Capacity** | High | Moderate | Low |
| **Onboard Processing** | Limited | Moderate | High |
| **Atmospheric Impact** | Minimal | Low | High |
| **Deployment Speed** | Slow | Medium | Fast |
| **Cost Structure** | High (low per area) | Medium | Low (high per area) |

amazon.com.au

All | drone

Hello, Sign in Account & Lists | Returns & Orders | Try Prime | 0 Cart

Hello Select your address

Best Sellers | Today's Deals | New Releases | Books | Gift Ideas | Electronics | Customer Service | Home | Computers

audible Free audiobook with trial

1-48 of over 60,000 results for "drone"

Sort by: Featured

**Amazon Prime**

- ☐ prime | Ships from Australia
- ☐ prime | International Shipping

**Avg. Customer Review**

- ★★★★☆ & Up
- ★★★☆☆ & Up
- ★★☆☆☆ & Up
- ★☆☆☆☆ & Up

**Department**

- Toys & Games
  - Hobby RC Drones & Multirotors
  - Toy Remote Control & Play Vehicles
- Electronics
  - Quadcopters & Accessories
  - Quadcopter Accessories
- Apps & Games
  - Game Apps
- See All 9 Departments

**Brand**

- ☐ DJI

Price and other details may vary based on size and colour

Amazon's Choice

DJI Mavic Mini - Drone FlyCam Quadcopter UAV with 2.7K Camera 3-Axis Gimbal GPS 30min Flight Time, less than 249g, Grey

★★★★½ 773

$597.93

prime Get it by Wednesday, September

FREE Delivery by Amazon

More Buying Choices

$596.00 (10 new offers)


REMOKING RC Drone with 720P FPV Wi-Fi HD Camera Live Video Racing Quadcopter Headless Mode 2.4GHz 360°flip 4 Channels Altitude Hold...

★★★★☆ 32

Drone with 4K Camera Live Video,EACHINE E520 WiFi FPV Drone for Adults with 4K HD Wide Angle Camera 1200Mah Long Flight time...

★★★★☆ 210

$169.40

prime Get it Tuesday, September 29 - Thursday, October 1

$13.53 shipping

Ages: 12 years and up

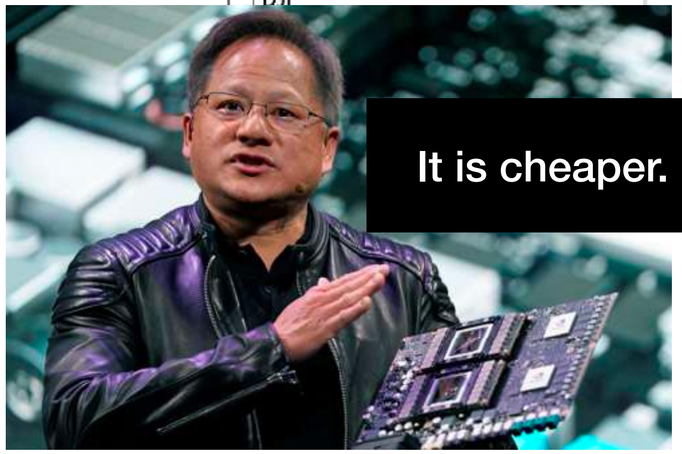
DJI Mavic 2 Pro - Drone Quadcopter UAV with Hasselblad Camera 3-Axis Gimbal HDR 4K Video Adjustable Aperture 20MP 1" CMOS Sensor, up...

★★★★½ 81

$2,249.10

prime Get it by Wednesday, September 9

FREE Delivery by Amazon



# Use Cases: What can Drones do? Why we study?

Drone is a new **aerial** platform.

- Accurate Delivery (e.g., send mask)
- Agriculture (e.g., pesticide)
- Event Detection (e.g. traffic jam)
- ….

https://www.youtube.com/watch?v=j65tqT-sQMo

# Outline

- University Dataset (ACM MM 2020)
- University-WX (PR 2024)
- GeoText-1652 (ECCV 2024)

# University-1652: A Multi-view Multi-source Benchmark for Drone-based Geo-localization

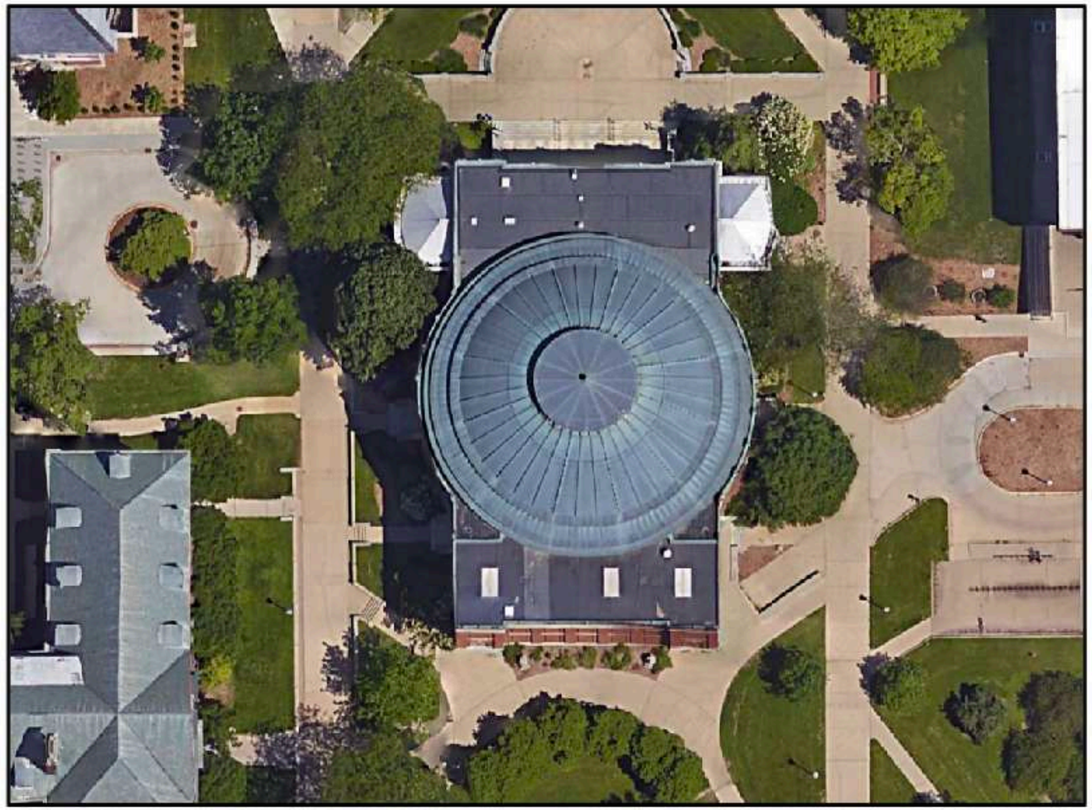
**Zhedong Zheng, Yunchao Wei, Yi Yang**

**ACM Multimedia**

# University-1652

- We consider one conventional task: cross-view Geo-localization.

**Ground-view Images** Gap **Satellite-view Images (GPS tag)**



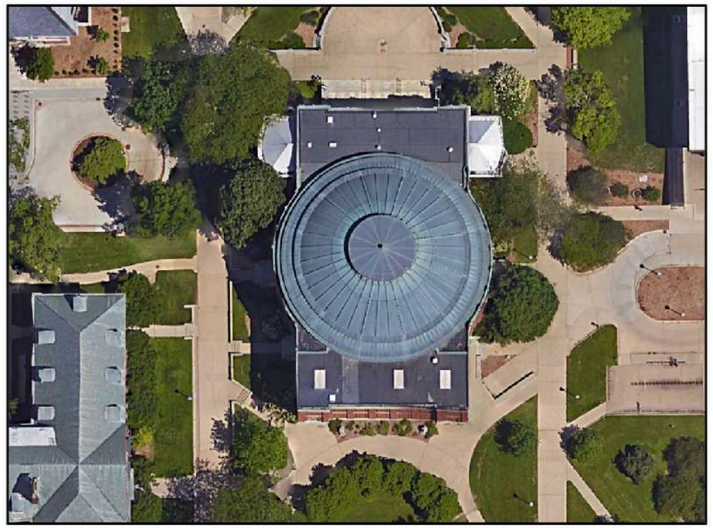
Limited Roof

Whole Roof

# We notice that the drone can be **a bridge.**

**Ground-view** ⟷ **Drone-view** ⟷ **Satellite-view (GPS tag)**

**No dataset to verify it.**



# University-1652

- We collect the data from three platforms of 1652 buildings.
- More training images per class (instead of image pairs).
- More viewpoints -> More intra-class variants

| Datasets | University-1652 | CVUSA [34] | CVACT [16] |
|---|---|---|---|
| #training | 701 × 71.64 | 35.5k × 2 | 35.5k × 2 |
| Platform | Drone, Ground, Satellite | Ground, Satellite | Ground, Satellite |
| #imgs./location | 54 + 16.64 + 1 | 1 + 1 | 1+1 |
| Target | Building | User | User |
| GeoTag | ✓ | ✓ | ✓ |
| Evaluation | Recall@K & AP | Recall@K | Recall@K |

- Me: I want to build one dataset.
- Supervisor: No! Too much cost.
- Me: We use free data from Internet.
- Supervisor: Do it!

# Building names from Wikipedia

| Building Names | |
|---|---|
| BibliothÃĺque Saint-Jean, University of Alberta | Clare Drake Arena |
| Foote Field | Myer Horowitz Theatre |
| National Institute for Nanotechnology | St Joseph's College, Edmonton |
| Stollery Children's Hospital | Universiade Pavilion, University of Alberta |
| University of Alberta Hospital | Alberta B. Farrington Softball Stadium |
| Decision Theater, University of Alberta | Gammage Memorial Auditorium |
| HarringtonâĂŞBirchett House | Industrial Arts Building |
| Irish Field | Louise Lincoln Kerr House and Studio |
| Matthews Hall, University of Alberta | Mona Plummer Aquatic Center |
| Old Main (Arizona State University) | Packard Stadium, University of Alberta |
| Security Building (Phoenix, Arizona) | Sun Devil Gym, University of Alberta |
| Sun Devil Stadium, University of Alberta | United States Post Office (Phoenix, Arizona) |
| Wells Fargo Arena (Tempe, Arizona) | Administration Building, University of Alberta |
| Wheeler Hall, University of Alberta | Marting Hall, University of Alberta |
| Malicky Center, University of Alberta | Burrell Memorial Observatory |
| Kleist Center for Art and Drama | Wilker Hall, University of Alberta |
| Kamm Hall, University of Alberta | Dietsch Hall, University of Alberta |
| Telfer Hall, University of Alberta | Ward Hall, University of Alberta |
| Thomas Center for Innovation and Growth (CIG) | Kulas Musical Arts Building, Baldwin Wallace University |
| Boesel Musical Arts Center, Baldwin Wallace University | Merner-Pfeiffer Hall, Baldwin Wallace University |
| Ritter Library, Baldwin Wallace University | Lindsay-Crossman Chapel, Baldwin Wallace University |
| Presidents House, Baldwin Wallace University | Student Activities Center (SAC), Baldwin Wallace University |
| Strosacker Hall (Union), Baldwin Wallace University | Bonds Hall, Baldwin Wallace University |
| Durst Welcome Center, Baldwin Wallace University | Lou Higgins Center, Baldwin Wallace University |
| Tressel Field @ Finnie Stadium, Baldwin Wallace University | Rutherford Library |
| Rudolph Ursprung Gymnasium, Baldwin Wallace University | Packard Athletic Center (formerly Bagley Hall), Baldwin Wallace University |
| Baldwin-Wallace College North Campus Historic District | Baldwin-Wallace College South Campus Historic District |
| Binghamton University Events Center, Binghamton University | Commonwealth Avenue, Boston University |
| Boston University Photonics Center, Boston University | Boston University School of Law, Boston University |
| Boston University Track and Tennis Center, Boston University | Boston University West Campus |

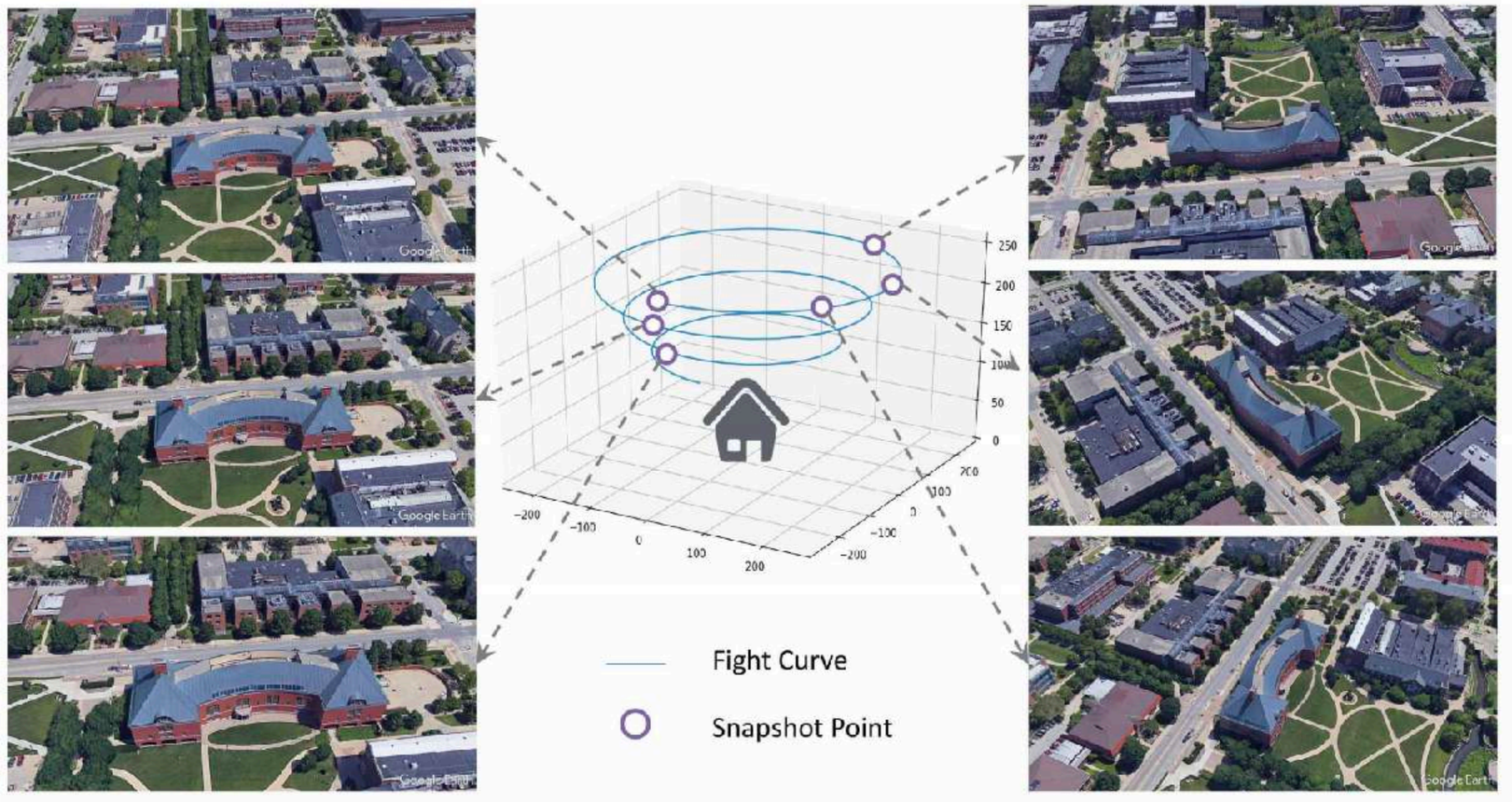
# Get latitude/longitude from GoogleMap

| Attributes | Value |
|---|---|
| name | Grainger Engineering Library |
| longitude | -88.22691719995214 |
| latitude | 40.11249969950067 |
| altitude | 18.56522342850079 |

| Attributes | Value |
|---|---|
| name | Foellinger Auditorium |
| longitude | -88.22728640012006 |
| latitude | 40.10594310015922 |
| altitude | 23.78598631063875 |

# 1. Drone-view Data

- Due to the privacy concerns and the cost, we deploy the simulated data via Google Earth. We write scripts to drive the engine as drone camera.




Google Earth

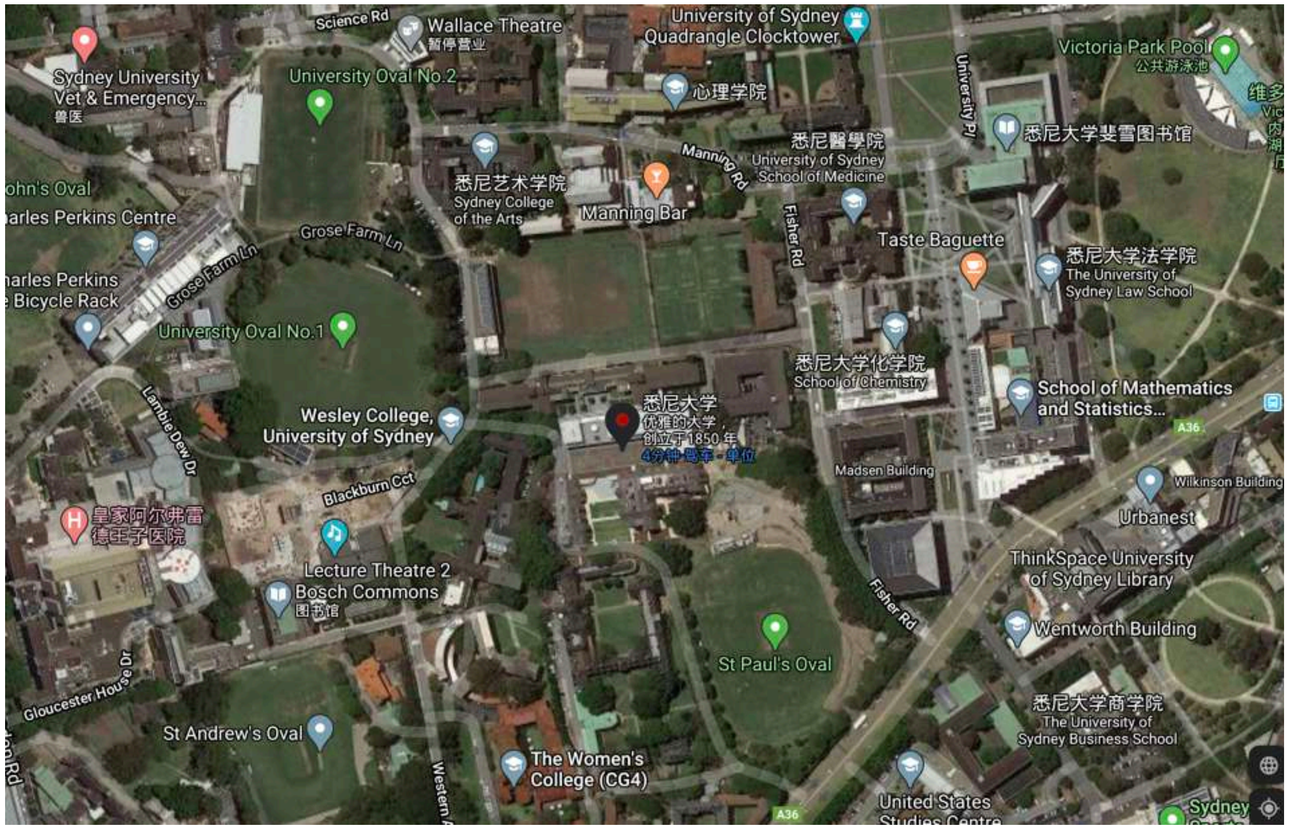

Google Earth


Google Earth


Google Earth

# 2. Ground-view Data from GoogleMap

# 3. Satellite-view Data from GoogleMap




# 4. Noisy Ground-view Data from GoogleImage

| Building Names | |
|---|---|
| BibliothÃĺque Saint-Jean, University of Alberta | Clare Drake Arena |
| Foote Field | Myer Horowitz Theatre |
| National Institute for Nanotechnology | St Joseph's College, Edmonton |
| Stollery Children's Hospital | Universiade Pavilion, University of Alberta |
| University of Alberta Hospital | Alberta B. Farrington Softball Stadium |
| Decision Theater, University of Alberta | Gammage Memorial Auditorium |
| HarringtonâĂŞBirchett House | Industrial Arts Building |
| Irish Field | Louise Lincoln Kerr House and Studio |
| Matthews Hall, University of Alberta | Mona Plummer Aquatic Center |
| Old Main (Arizona State University) | Packard Stadium, University of Alberta |
| Security Building (Phoenix, Arizona) | Sun Devil Gym, University of Alberta |
| Sun Devil Stadium, University of Alberta | United States Post Office (Phoenix, Arizona) |
| Wells Fargo Arena (Tempe, Arizona) | Administration Building, University of Alberta |
| Wheeler Hall, University of Alberta | Marting Hall, University of Alberta |
| Malicky Center, University of Alberta | Burrell Memorial Observatory |
| Kleist Center for Art and Drama | Wilker Hall, University of Alberta |
| Kamm Hall, University of Alberta | Dietsch Hall, University of Alberta |
| Telfer Hall, University of Alberta | Ward Hall, University of Alberta |
| Thomas Center for Innovation and Growth (CIG) | Kulas Musical Arts Building, Baldwin Wallace University |
| Boesel Musical Arts Center, Baldwin Wallace University | Merner-Pfeiffer Hall, Baldwin Wallace University |
| Ritter Library, Baldwin Wallace University | Lindsay-Crossman Chapel, Baldwin Wallace University |
| Presidents House, Baldwin Wallace University | Student Activities Center (SAC), Baldwin Wallace University |
| Strosacker Hall (Union), Baldwin Wallace University | Bonds Hall, Baldwin Wallace University |
| Durst Welcome Center, Baldwin Wallace University | Lou Higgins Center, Baldwin Wallace University |
| Tressel Field @ Finnie Stadium, Baldwin Wallace University | Rutherford Library |
| Rudolph Ursprung Gymnasium, Baldwin Wallace University | Packard Athletic Center (formerly Bagley Hall), Baldwin Wallace University |
| Baldwin-Wallace College North Campus Historic District | Baldwin-Wallace College South Campus Historic District |
| Binghamton University Events Center, Binghamton University | Commonwealth Avenue, Boston University |
| Boston University Photonics Center, Boston University | Boston University School of Law, Boston University |
| Boston University Track and Tennis Center, Boston University | Boston University West Campus |

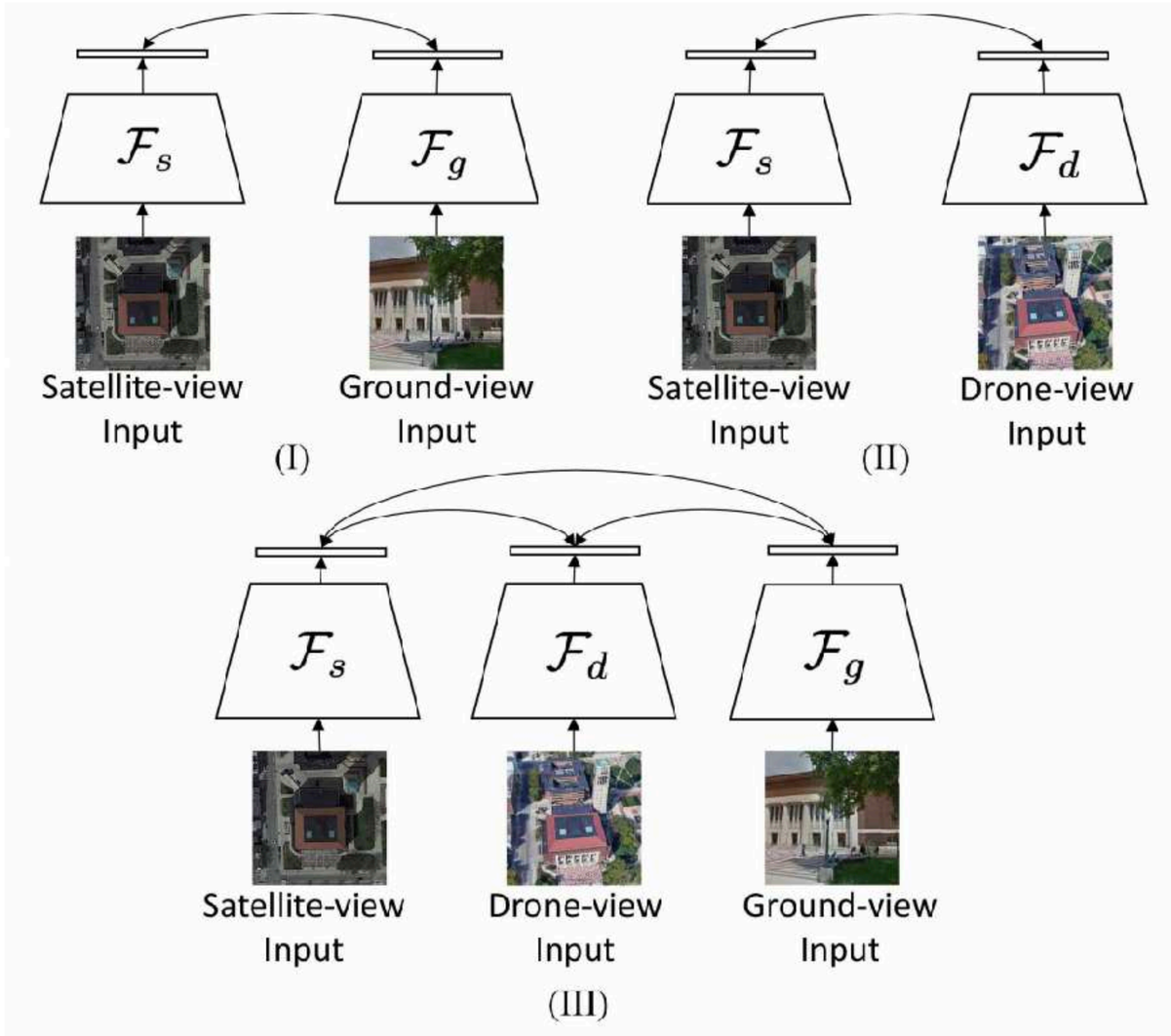
- We search the building name and download images from GoogleImage
- We then remove the indoor images and duplicate images.

**Krause J, Sapp B, Howard A, et al. The unreasonable effectiveness of noisy data for fine-grained recognition[C]//European Conference on Computer Vision. Springer, Cham, 2016: 301-320.**

# Baseline

Flexible and Strong Baseline

- Objective: Instance Loss (Share Classifier)
- Structure: Generally, the backbone
- network do not share low-level patterns




New data ->

add one branch!

Zheng Z, Zheng L, Garrett M, et al. Dual-Path Convolutional Image-Text Embeddings with Instance Loss[J]. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 2020, 16(2): 1-23.

# Baseline

## CVUSA

| Methods | R@1 | R@5 | R@10 | R@Top1% |
|---|---|---|---|---|
| Workman [31] ICCV 2015 | - | - | - | 34.40 |
| Zhai [34] CVPR 2017 | - | - | - | 43.20 |
| Vo [29] ECCV 2016 | - | - | - | 63.70 |
| CVM-Net [11] CVPR 2018 | 18.80 | 44.42 | 57.47 | 91.54 |
| Orientation [16]† CVPR 2019 | 27.15 | 54.66 | 67.54 | **93.91** |
| Ours | **43.91** | **66.38** | **74.58** | 91.78 |

**Table 9: Comparison of results on the two-view dataset CVUSA [34]. †: The method utilizes extra orientation information as input.**

## Oxford and Paris

| Method | Oxford | Paris | ROxf (M) | RPar (M) | ROxf (H) | RPar (H) |
|---|---|---|---|---|---|---|
| ImageNet | 3.30 | 6.77 | 4.17 | 8.20 | 2.09 | 4.24 |
| $\mathcal{F}_s$ | 9.24 | 13.74 | 5.83 | 13.79 | 2.08 | 6.40 |
| $\mathcal{F}_g$ | 25.80 | 28.77 | 15.52 | 24.24 | 3.69 | 10.29 |

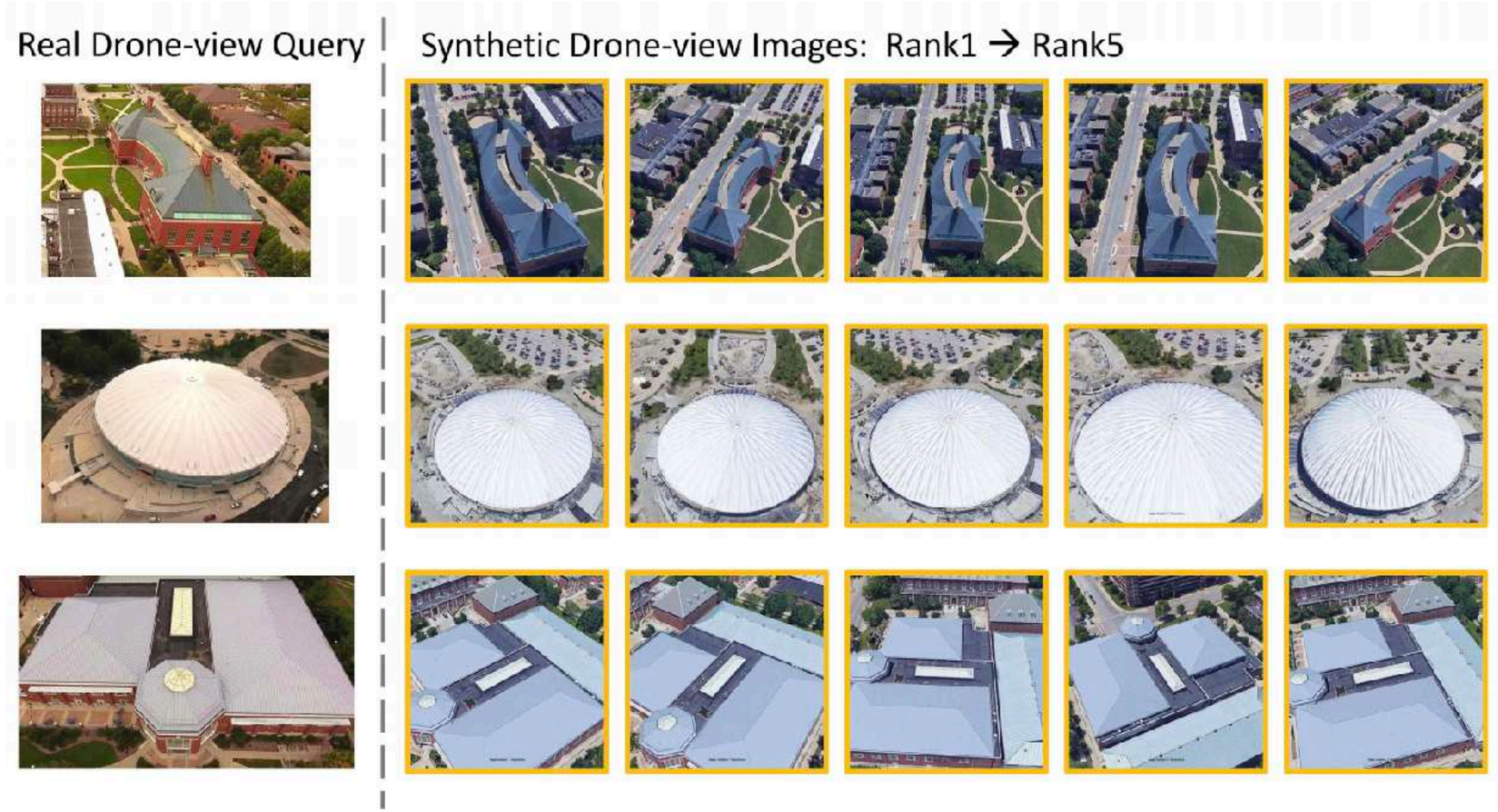
**Table 10: Transfer learning from University-1652 to small-scale datasets. We show the AP (%) accuracy on Oxford [19], Paris [20], ROxford and RParis [21]. For ROxford and RParis, we report results in both medium (M) and hard (H) settings.**

# Ground-view query vs. drone-view query.

| Query → Gallery | R@1 | R@5 | R@10 | AP |
|---|---|---|---|---|
| Ground → Satellite | 1.20 | 4.61 | 7.56 | 2.52 |
| Drone → Satellite | 58.49 | 78.67 | 85.23 | 63.13 |
| *m*Ground → Satellite | 1.71 | 6.56 | 10.98 | 3.33 |
| *m*Drone → Satellite | 69.33 | 86.73 | 91.16 | 73.14 |

**Table 4: Ground-view query vs. drone-view query. *m* denotes multiple-query setting. The result suggests that drone-view images are superior to ground-view images when retrieving satellite-view images.**

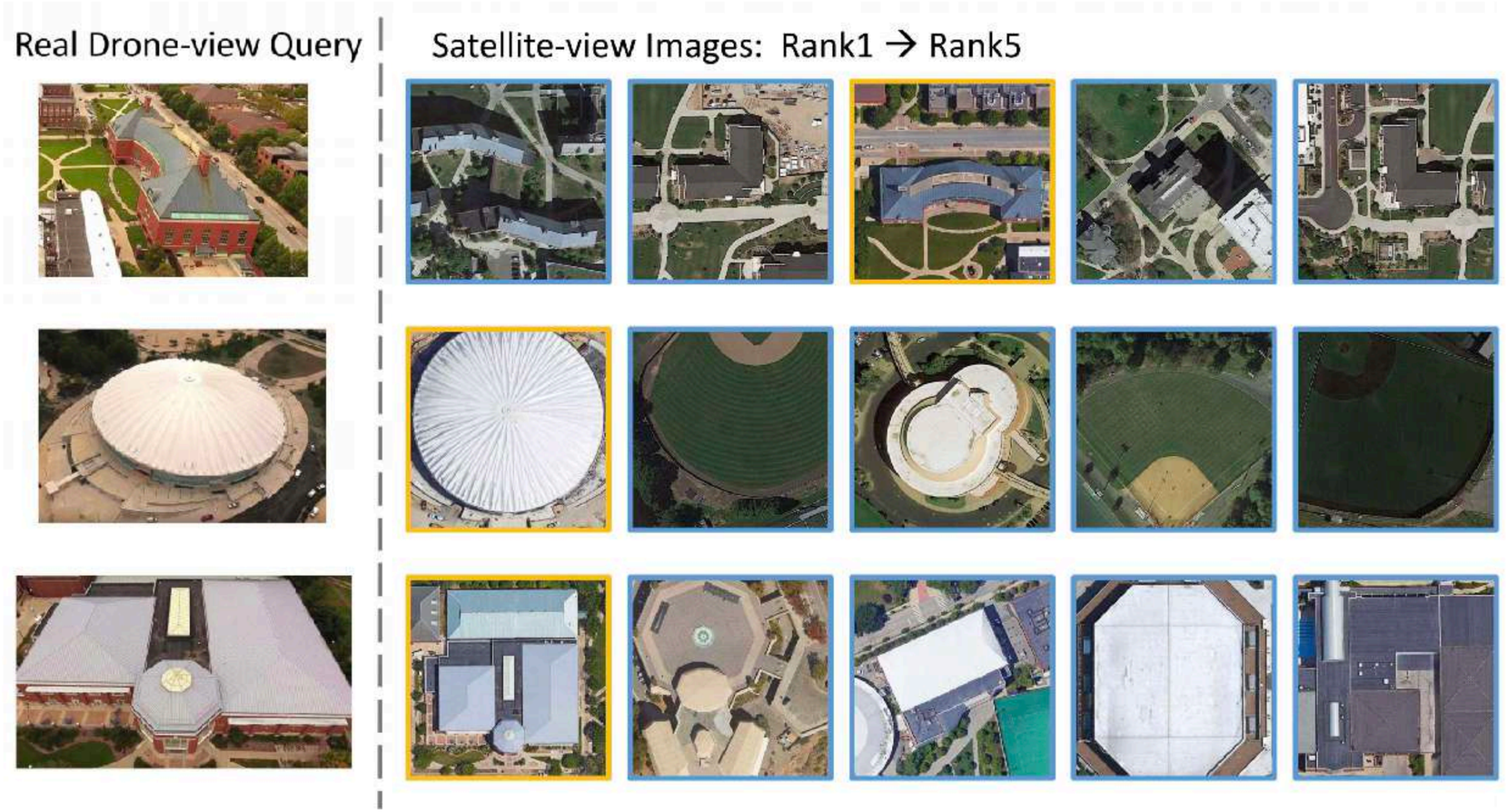
# Apply the model trained on University-1652 to real drone videos.




[Fly High #1 "UIUC" https://www.youtube.com/watch?v=jOC-WJW7GAg](https://www.youtube.com/watch?v=jOC-WJW7GAg)

The model haven't seen any data of UIUC.

# Apply the model trained on University-1652 to real drone videos.




The model haven’t  seen any data of UIUC.

**Scalability**

# Ablation Studies

## Different Loss Functions

| Loss | Drone → Satellite | | Satellite → Drone | |
|---|---|---|---|---|
| | R@1 | AP | R@1 | AP |
| Contrastive Loss | 52.39 | 57.44 | 63.91 | 52.24 |
| Triplet Loss (margin=0.3) | 55.18 | 59.97 | 63.62 | 53.85 |
| Triplet Loss (margin=0.5) | 53.58 | 58.60 | 64.48 | 53.15 |
| Weighted Soft Margin Triplet Loss | 53.21 | 58.03 | 65.62 | 54.47 |
| Instance Loss | 58.23 | 62.91 | 74.47 | 59.45 |

**Table 5: Ablation study of different loss terms. To fairly compare the five loss terms, we trained the five models on satellite-view and drone-view data, and hold out the ground-view data. For contrastive loss, triplet loss and weighted soft margin triplet loss, we also apply the hard-negative sampling policy.**

## Whether Share Weights

| Method | Drone → Satellite | | Satellite → Drone | |
|---|---|---|---|---|
| | R@1 | AP | R@1 | AP |
| Not sharing weights | 39.84 | 45.91 | 50.36 | 40.71 |
| Sharing weights | 58.49 | 63.31 | 71.18 | 58.74 |

**Table 6: Ablation study. With/without sharing CNN weights on University-1652. The result suggests that sharing weights could help to regularize the CNN model.**

## Different Input Sizes

| Image Size | Drone → Satellite | | Satellite → Drone | |
|---|---|---|---|---|
| | R@1 | AP | R@1 | AP |
| 256 | 58.49 | 63.31 | 71.18 | 58.74 |
| 384 | 62.99 | 67.69 | 75.75 | 62.09 |
| 512 | 59.69 | 64.80 | 73.18 | 59.40 |

**Table 7: Ablation study of different input sizes on the University-1652 dataset.**

# Future Works - Boost Performance

We  run a leaderboard.
You are welcomed to push the state-of-the-art performance.

## Awesome Geo-localization

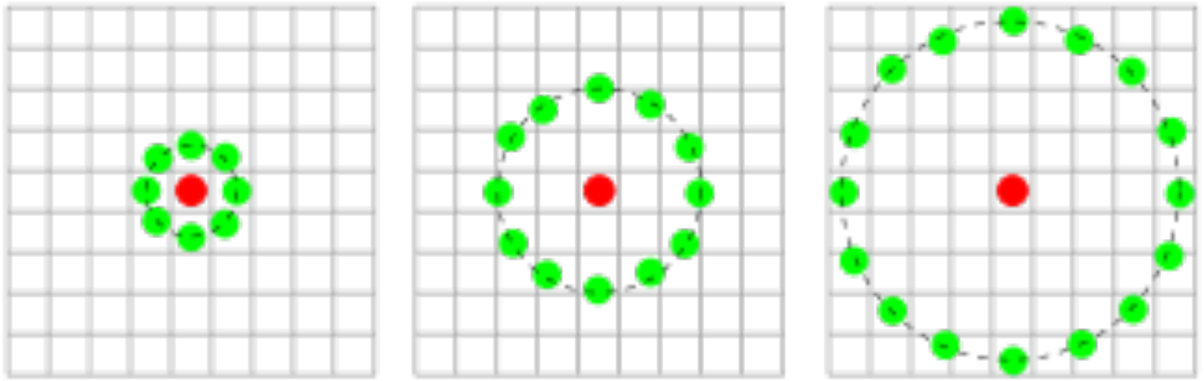
### University-1652

| Methods | R@1 | AP | R@1 | AP | Reference |
|---|---|---|---|---|---|
| Contrastive Loss | 52.39 | 57.44 | 63.91 | 52.24 | |
| Triplet Loss (margin=0.3) | 55.18 | 59.97 | 63.62 | 53.85 | |
| Triplet Loss (margin=0.5) | 53.58 | 58.60 | 64.48 | 53.15 | |
| Weighted Soft Margin Triplet Loss | 53.21 | 58.03 | 65.62 | 54.47 | |
| Instance Loss | 58.23 | 62.91 | 74.47 | 59.45 | |

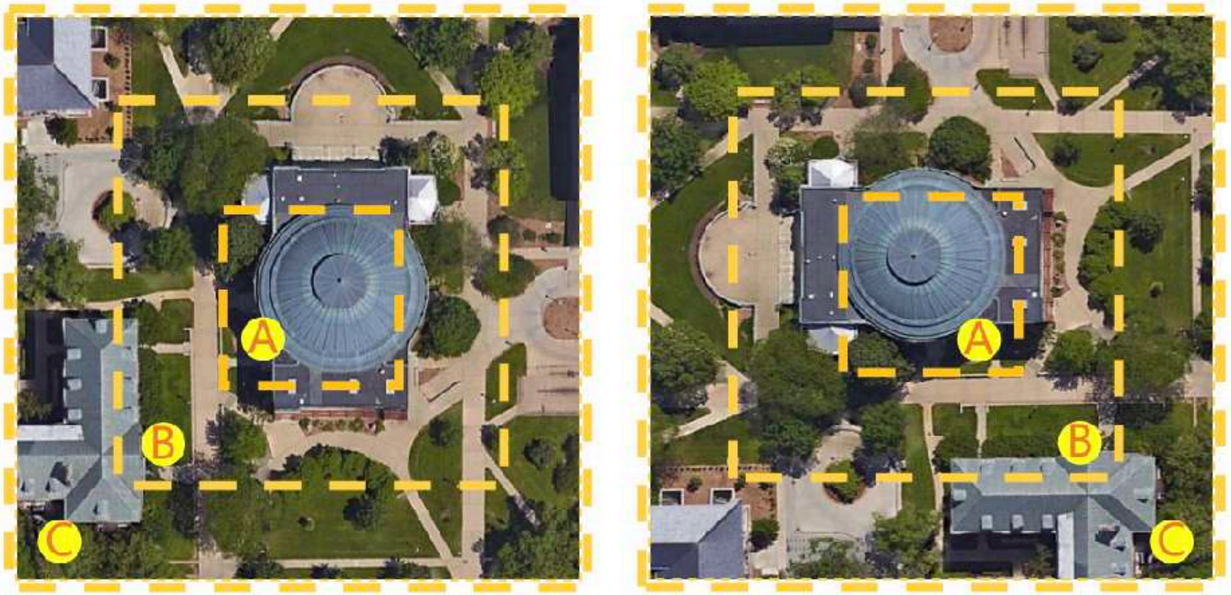
TIP 2022
ICCV 2023
TGRS 2023
TGRS 2024
ACM MM 2024

https://github.com/layumi/University1652-Baseline/tree/master/State-of-the-art

# Extension - LBP



Local binary patterns (LBP) descriptor

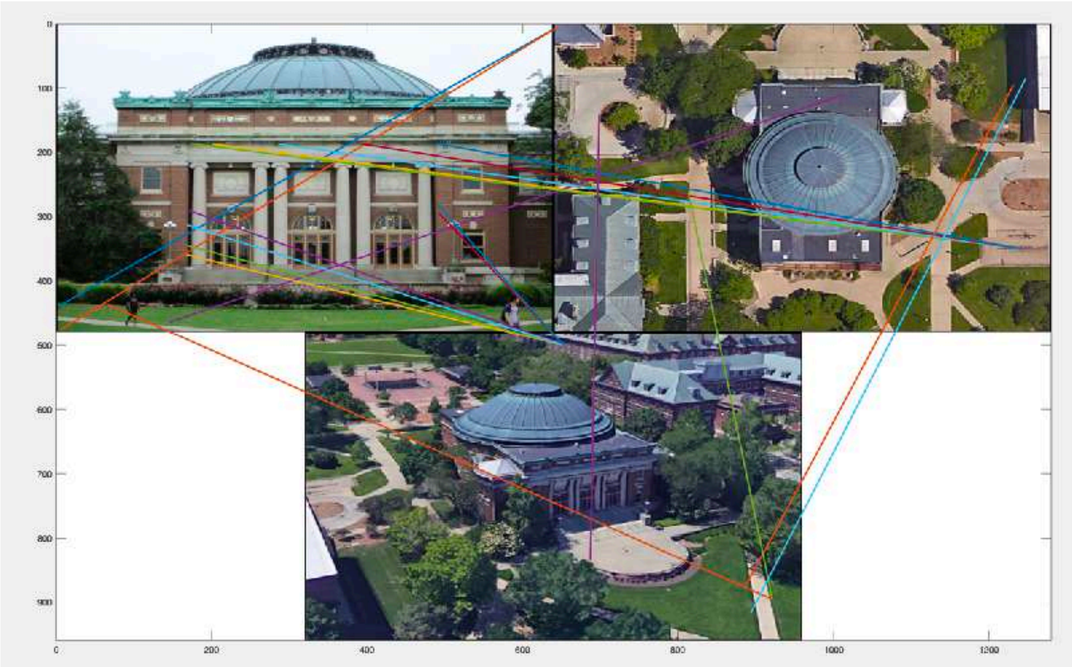



LPN

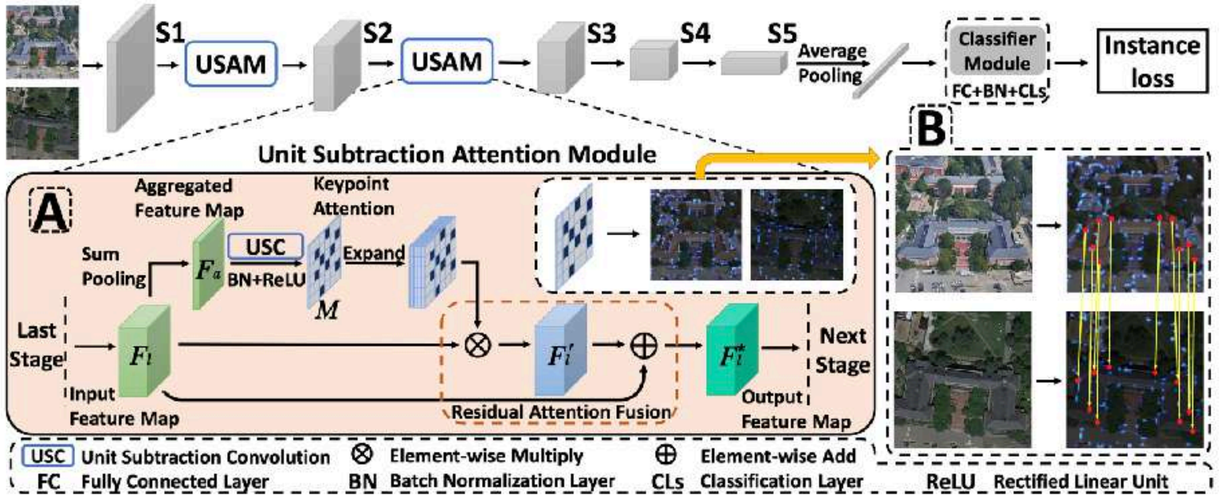
Tingyu W, Zhedong Z, Chenggang Y, and Yi, Y. Each Part Matters: Local Patterns Facilitate Cross-view Geo-localization. TCSVT 2021.

# Extension - Keypoint Matching

## SIFT does not work very well. Deeply-learned Methods are needed.

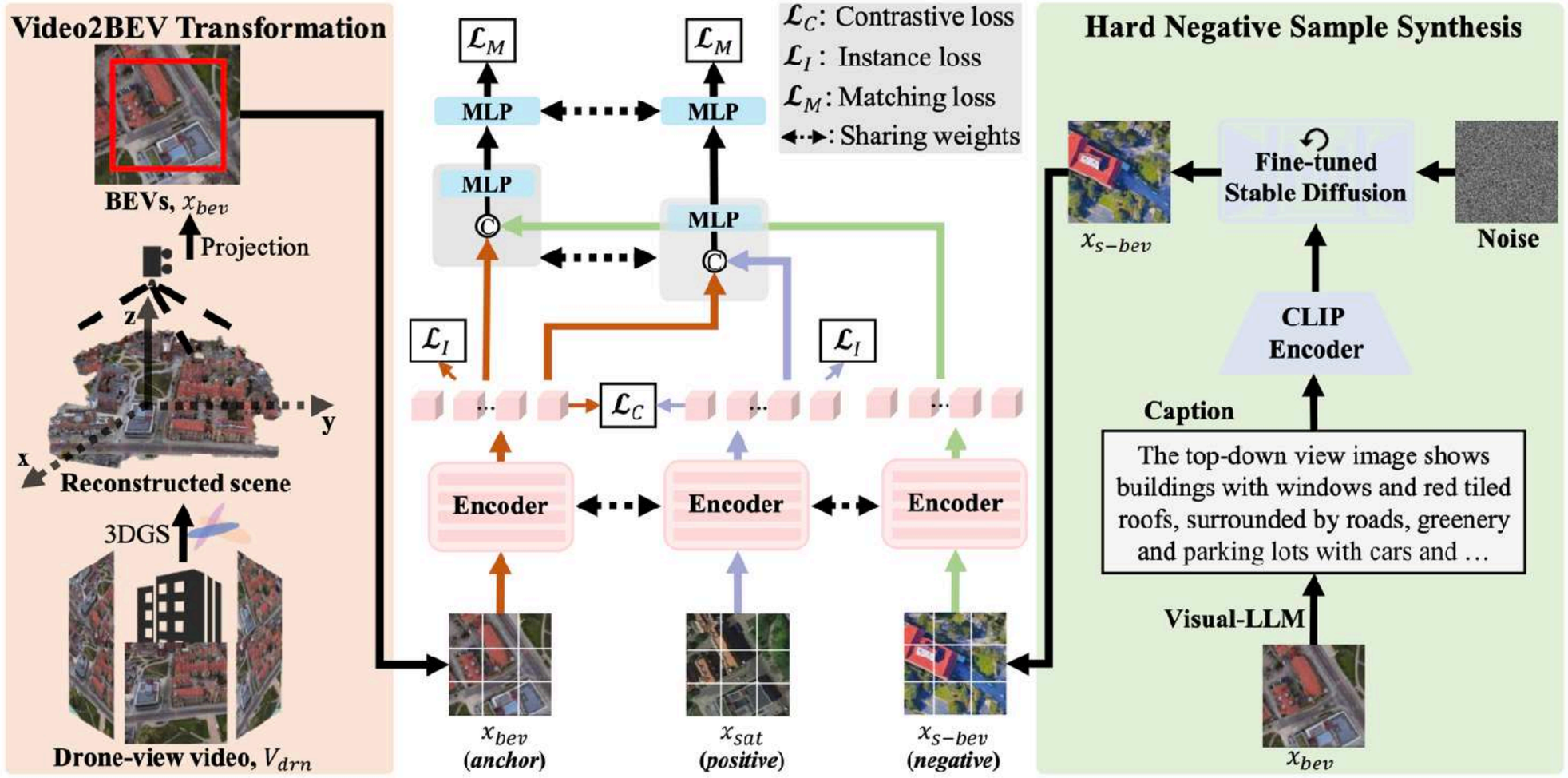


## We design a differentiable key point detection.




Lin J, Zheng Z, Zhong Z, et al. Joint representation learning and keypoint detection for cross-view geo-localization[J]. IEEE Transactions on Image Processing, 2022, 31: 3780-3792.

# Extension - 3D




Ju, H., Huang, S., Liu, S., & Zheng, Z. (2024). Video2bev: Transforming drone videos to bevs for video-based geo-localization. *arXiv:2411.13610*.

# Data License

- We carefully check the data license from Google. There are two main points.
- First, the data of Google Map and Google Earth could be used based on fair usage. We follow the guideline on this official website 3 .
- Second, several existing datasets have utilized the Google data. In practice, we adopt a similar policy of existing datasets 4, 5 to release the dataset based on the academic request.

3. https://www.google.com/permissions/geoguidelines/
4. http://www.ok.ctrl.titech.ac.jp/~torii/project/247/
5. http://mvrl.cs.uky.edu/datasets/cvusa/

# Outline

- University Dataset (ACM MM 2020)
- University-WX (PR 2024)
- GeoText-1652 (ECCV 2024)

# Multiple-environment Self-adaptive Network for Aerial-view Geo-localization

**Tingyu Wang, Zhedong Zheng, Yaoqi Sun, Chenggang Yan, Yi Yang, Tat-Seng Chua**

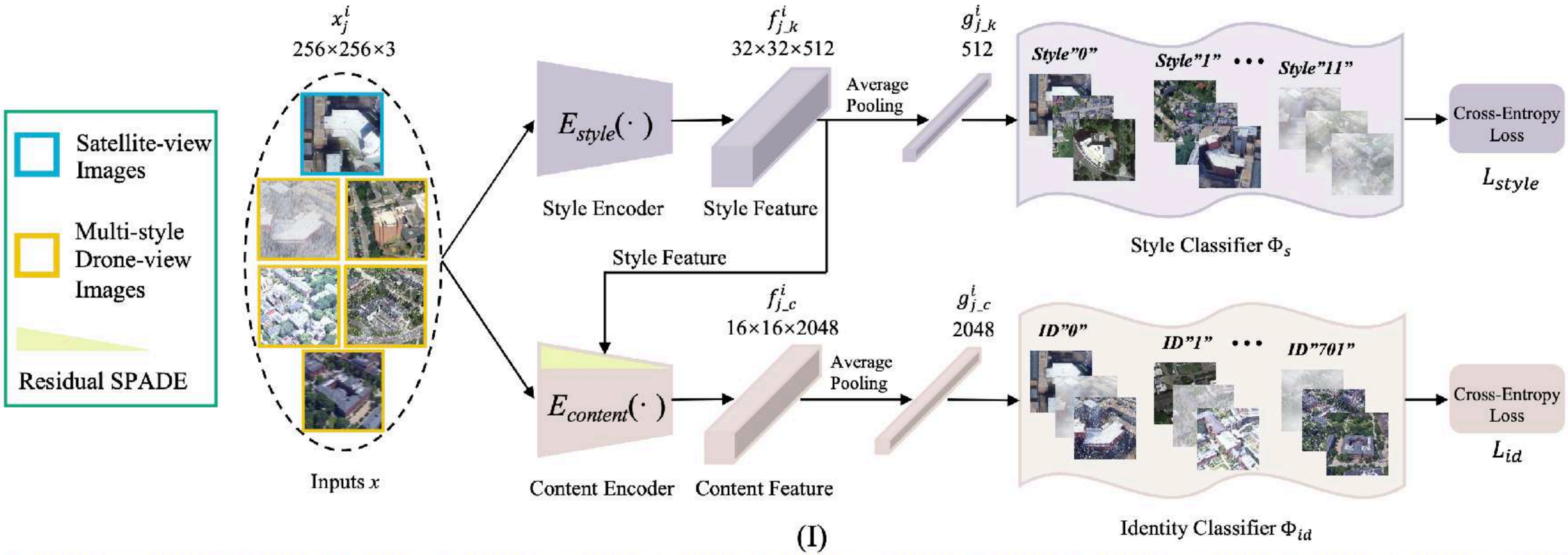
Pattern Recognition

# University-1652 meets Extreme Weather.

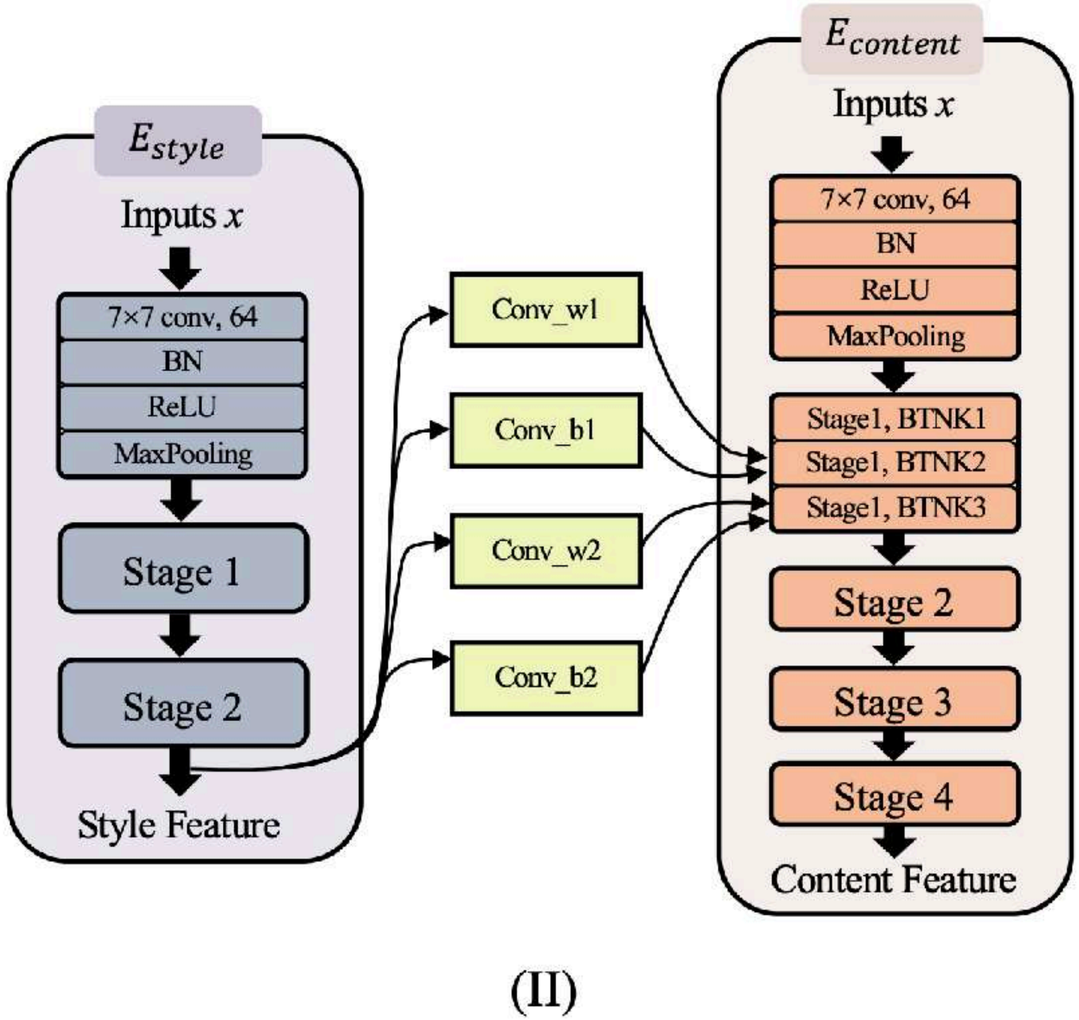
# Extreme Weather compromise the performance.

| Method | Normal | | Fog | | Rain | | Snow | | Fog +Rain | | Fog +Snow | | Rain +Snow | | Dark | | Over -exposure | | Wind | | Mean ↑ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | R@1 | AP | R@1 | AP | R@1 | AP | R@1 | AP | R@1 | AP | R@1 | AP | R@1 | AP | R@1 | AP | R@1 | AP | R@1 | AP | R@1 | AP |
| **Drone → Satellite** | | | | | | | | | | | | | | | | | | | | | | |
| VGG16 [53] | 59.98 | 64.69 | 56.21 | 61.11 | 53.97 | 58.90 | 50.07 | 55.08 | 50.43 | 55.63 | 42.77 | 48.01 | 51.08 | 56.10 | 39.10 | 44.30 | 45.16 | 50.47 | 50.84 | 56.05 | 49.96 | 55.03 |
| Zheng *et al.* [1] | 67.83 | 71.74 | 60.97 | 65.23 | 60.29 | 64.61 | 55.58 | 60.09 | 54.75 | 59.40 | 44.35 | 49.78 | 57.61 | 62.03 | 39.70 | 44.65 | 51.85 | 56.75 | 58.28 | 62.83 | 55.17 | 59.71 |
| ResNet-101 [21] | 70.07 | 73.04 | 63.87 | 68.22 | 63.34 | 67.59 | 59.75 | 64.15 | 57.45 | 62.12 | 48.31 | 53.28 | 60.25 | 64.68 | 46.12 | 51.02 | 56.34 | 61.23 | 62.13 | 66.63 | 58.76 | 63.29 |
| DenseNet121 [20] | 69.48 | 73.26 | 64.25 | 68.47 | 63.47 | 67.64 | 59.29 | 63.70 | 59.68 | 64.13 | 50.41 | 55.20 | 60.21 | 64.57 | 48.57 | 53.41 | 54.04 | 58.88 | 60.74 | 65.14 | 59.01 | 63.44 |
| Swin-T [54] | 69.27 | 73.18 | 66.46 | 70.52 | 65.44 | 69.60 | 61.79 | 66.23 | 63.96 | 68.21 | **56.44** | **61.07** | 62.68 | 67.02 | 50.27 | 55.18 | 55.46 | 60.29 | 63.81 | 68.17 | 61.56 | 65.95 |
| IBN-Net [18] | 72.35 | 75.85 | 66.68 | 70.64 | 67.95 | 71.73 | 62.77 | 66.85 | 62.64 | 66.84 | 51.09 | 55.79 | 64.07 | 68.13 | 50.72 | 55.53 | 57.97 | 62.52 | 66.73 | 70.68 | 62.30 | 66.46 |
| LPN [2] | 74.33 | 77.60 | 69.31 | 72.95 | 67.96 | 71.72 | 64.90 | 68.85 | 64.51 | 68.52 | 54.16 | 58.73 | 65.38 | 69.29 | 53.68 | 58.10 | 60.90 | 65.27 | 66.46 | 70.35 | 64.16 | 68.14 |
| Ours | **74.48** | **77.83** | **69.47** | **73.24** | **70.55** | **74.14** | **65.72** | **69.70** | **65.59** | **69.64** | 54.69 | 59.24 | **66.64** | **70.55** | **53.85** | **58.49** | **61.05** | **65.51** | **69.45** | **73.22** | **65.15** | **69.16** |
| **Satellite → Drone** | | | | | | | | | | | | | | | | | | | | | | |
| VGG16 [53] | 75.89 | 58.50 | 75.18 | 55.42 | 71.61 | 53.03 | 68.19 | 48.29 | 71.18 | 49.34 | 65.48 | 40.87 | 69.47 | 50.03 | 64.34 | 35.74 | 64.91 | 44.20 | 68.90 | 49.53 | 69.52 | 48.50 |
| Zheng *et al.* [1] | 83.45 | 67.94 | 79.60 | 61.12 | 77.60 | 59.73 | 73.18 | 55.07 | 75.89 | 54.45 | 70.76 | 43.26 | 74.75 | 56.44 | 69.47 | 39.25 | 72.18 | 51.91 | 76.46 | 57.59 | 75.33 | 54.68 |
| ResNet-101 [21] | 85.73 | 71.79 | 82.45 | 66.46 | 81.46 | 65.68 | 79.74 | 61.72 | 79.74 | 60.59 | 74.75 | 50.31 | 80.17 | 62.61 | 75.32 | 45.37 | 79.60 | 58.21 | 82.31 | 64.67 | 80.13 | 60.74 |
| DenseNet121 [20] | 83.74 | 70.34 | 82.31 | 66.32 | 81.17 | 65.23 | 78.60 | 60.33 | 79.46 | 61.66 | 74.61 | 51.14 | 78.46 | 61.68 | 74.47 | 47.88 | 74.32 | 55.26 | 78.32 | 61.63 | 78.55 | 60.15 |
| Swin-T [54] | 80.74 | 68.94 | 81.03 | 67.46 | 81.17 | 66.39 | 78.46 | 61.33 | 79.17 | 64.65 | 74.89 | 56.57 | 78.89 | 63.49 | 75.61 | 48.43 | 76.60 | 56.57 | 78.74 | 64.45 | 78.53 | 61.83 |
| IBN-Net [18] | 86.31 | 73.54 | 84.59 | 67.61 | 84.74 | 69.03 | 80.88 | 64.44 | 83.31 | 63.71 | 77.89 | 52.14 | 83.02 | 65.74 | 78.46 | 50.77 | 79.46 | 58.64 | 84.02 | 67.94 | 82.27 | 63.36 |
| LPN [2] | 87.02 | **75.19** | 86.16 | **71.34** | 83.88 | 69.49 | 82.88 | 65.39 | 84.59 | 66.28 | 79.60 | **55.19** | 84.17 | 66.26 | **82.88** | 52.05 | 81.03 | **62.24** | 84.14 | 67.35 | 83.64 | 65.08 |
| Ours | **88.02** | 75.10 | **87.87** | 69.85 | **87.73** | **71.12** | **83.74** | **66.52** | **85.02** | **67.78** | **80.88** | 54.26 | **84.88** | **67.75** | 80.74 | **53.01** | **81.60** | 62.09 | **86.31** | **70.03** | **84.68** | **65.75** |

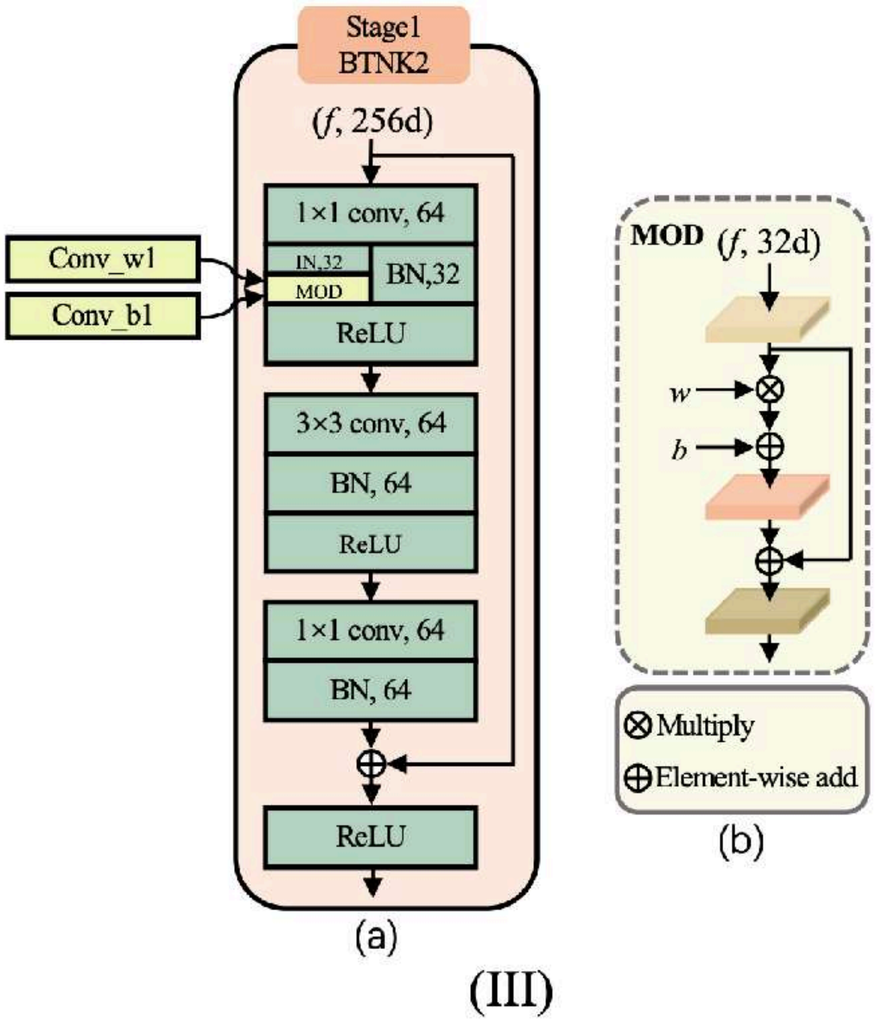
# Vanilla Method : Just Estimate Weather



# How to Remove?
# Replace W and B in Instance Norm




(II)




(a) (b)

(III)

# Extreme Weather remains Challenging.

| Method | Normal | | Fog | | Rain | | Snow | | Fog +Rain | | Fog +Snow | | Rain +Snow | | Dark | | Over -exposure | | Wind | | **Mean ↑** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | R@1 | AP | R@1 | AP | R@1 | AP | R@1 | AP | R@1 | AP | R@1 | AP | R@1 | AP | R@1 | AP | R@1 | AP | R@1 | AP | R@1 | AP |
| **Drone → Satellite** | | | | | | | | | | | | | | | | | | | | | | |
| VGG16 [56] | 59.98 | 64.69 | 56.21 | 61.11 | 53.97 | 58.90 | 50.07 | 55.08 | 50.43 | 55.63 | 42.77 | 48.01 | 51.08 | 56.10 | 39.10 | 44.30 | 45.16 | 50.47 | 50.84 | 56.05 | 49.96 | 55.03 |
| Zheng *et al.* [1] | 67.83 | 71.74 | 60.97 | 65.23 | 60.29 | 64.61 | 55.58 | 60.09 | 54.75 | 59.40 | 44.85 | 49.78 | 57.61 | 62.03 | 39.70 | 44.65 | 51.85 | 56.75 | 58.28 | 62.83 | 55.17 | 59.71 |
| ResNet-101 [21] | 70.07 | 73.04 | 63.87 | 68.22 | 63.34 | 67.59 | 59.75 | 64.15 | 57.45 | 62.12 | 48.31 | 53.28 | 60.25 | 64.68 | 46.12 | 51.02 | 56.34 | 61.23 | 62.13 | 66.63 | 58.76 | 63.29 |
| DenseNet121 [20] | 69.48 | 73.26 | 64.25 | 68.47 | 63.47 | 67.64 | 59.29 | 63.70 | 59.68 | 64.13 | 50.41 | 55.20 | 60.21 | 64.57 | 48.57 | 53.41 | 54.04 | 58.88 | 60.74 | 65.14 | 59.01 | 63.44 |
| Swin-T [57] | 69.27 | 73.18 | 66.46 | 70.52 | 65.44 | 69.60 | 61.79 | 66.23 | 63.96 | 68.21 | **56.44** | **61.07** | 62.68 | 67.02 | 50.27 | 55.18 | 55.46 | 60.29 | 63.81 | 68.17 | 61.56 | 65.95 |
| IBN-Net [18] | 72.35 | 75.85 | 66.68 | 70.64 | 67.95 | 71.73 | 62.77 | 66.85 | 62.64 | 66.84 | 51.09 | 55.79 | 64.07 | 68.13 | 50.72 | 55.53 | 57.97 | 62.52 | 66.73 | 70.68 | 62.30 | 66.46 |
| LPN [2] | 74.33 | 77.60 | [illegible] | [illegible] | 67.96 | 71.72 | 64.90 | 68.85 | 64.51 | 68.52 | 54.16 | 58.73 | 65.38 | 69.29 | 53.68 | 58.10 | 60.90 | 65.27 | 66.46 | 70.35 | 64.16 | 68.14 |
| Ours | **74.48** | **77.83** | **69.47** | **73.[illegible]** | **70.55** | **74.14** | **65.72** | **69.70** | **65.59** | **69.64** | 54.69 | 59.24 | **66.64** | **70.55** | **53.85** | **58.49** | **61.05** | **65.51** | **69.45** | **73.22** | **65.15** | **69.16** |
| **Satellite → Drone** | | | | | | | | | | | | | | | | | | | | | | |
| VGG16 [56] | 75.89 | 58.50 | 75.18 | 55.42 | 71.61 | 53.03 | 68.19 | 48.29 | 71.18 | 49.34 | 65.48 | 40.87 | 69.47 | 50.03 | 64.34 | 35.74 | 64.91 | 44.20 | 68.90 | 49.53 | 69.52 | 48.50 |
| Zheng *et al.* [1] | 83.45 | 67.94 | 79.60 | 61.12 | 77.60 | 59.73 | 73.18 | 55.07 | 75.89 | 54.45 | 70.76 | 43.26 | 74.75 | 56.44 | 69.47 | 39.25 | 72.18 | 51.91 | 76.46 | 57.59 | 75.33 | 54.68 |
| ResNet-101 [21] | 85.73 | 71.79 | 82.45 | 66.46 | 81.46 | 65.68 | 79.74 | 61.72 | 79.74 | 60.59 | 74.75 | 50.31 | 80.17 | 62.61 | 75.32 | 45.37 | 79.60 | 58.21 | 82.31 | 64.67 | 80.13 | 60.74 |
| DenseNet121 [20] | 83.74 | 70.34 | 82.31 | 66.32 | 81.17 | 65.23 | 78.60 | 60.33 | 79.46 | 61.66 | 74.61 | 51.14 | 78.46 | 61.68 | 74.47 | 47.88 | 74.32 | 55.26 | 78.32 | 61.63 | 78.55 | 60.15 |
| Swin-T [57] | 80.74 | 68.94 | 81.03 | 67.46 | 81.17 | 66.39 | 78.46 | 61.33 | 79.17 | 64.65 | 74.89 | 56.57 | 78.89 | 63.49 | 75.61 | 48.43 | 76.60 | 56.57 | 78.74 | 64.45 | 78.53 | 61.83 |
| IBN-Net [18] | 86.31 | 73.54 | 84.59 | 67.61 | 84.74 | 69.03 | 80.88 | 64.44 | 83.31 | 63.71 | 77.89 | 52.14 | 83.02 | 65.74 | 78.46 | 50.77 | 79.46 | 58.64 | 84.02 | 67.94 | 82.27 | 63.36 |
| LPN [2] | 87.02 | **75.19** | [illegible] | [illegible] | 83.88 | 69.49 | 82.88 | 65.39 | 84.59 | 66.28 | 79.60 | **55.19** | 84.17 | 66.26 | **82.88** | 52.05 | 81.03 | **62.24** | 84.14 | 67.35 | 83.64 | 65.08 |
| Ours | **88.02** | 75.10 | **87.87** | [illegible] | **87.73** | **71.12** | **83.74** | **66.52** | **85.02** | **67.78** | **80.88** | 54.26 | **84.88** | **67.75** | 80.74 | **53.01** | **81.60** | 62.09 | **86.31** | **70.03** | **84.68** | **65.75** |

# Outline

- University Dataset (ACM MM 2020)
- University-WX (PR 2024)
- GeoText-1652 (ECCV 2024)

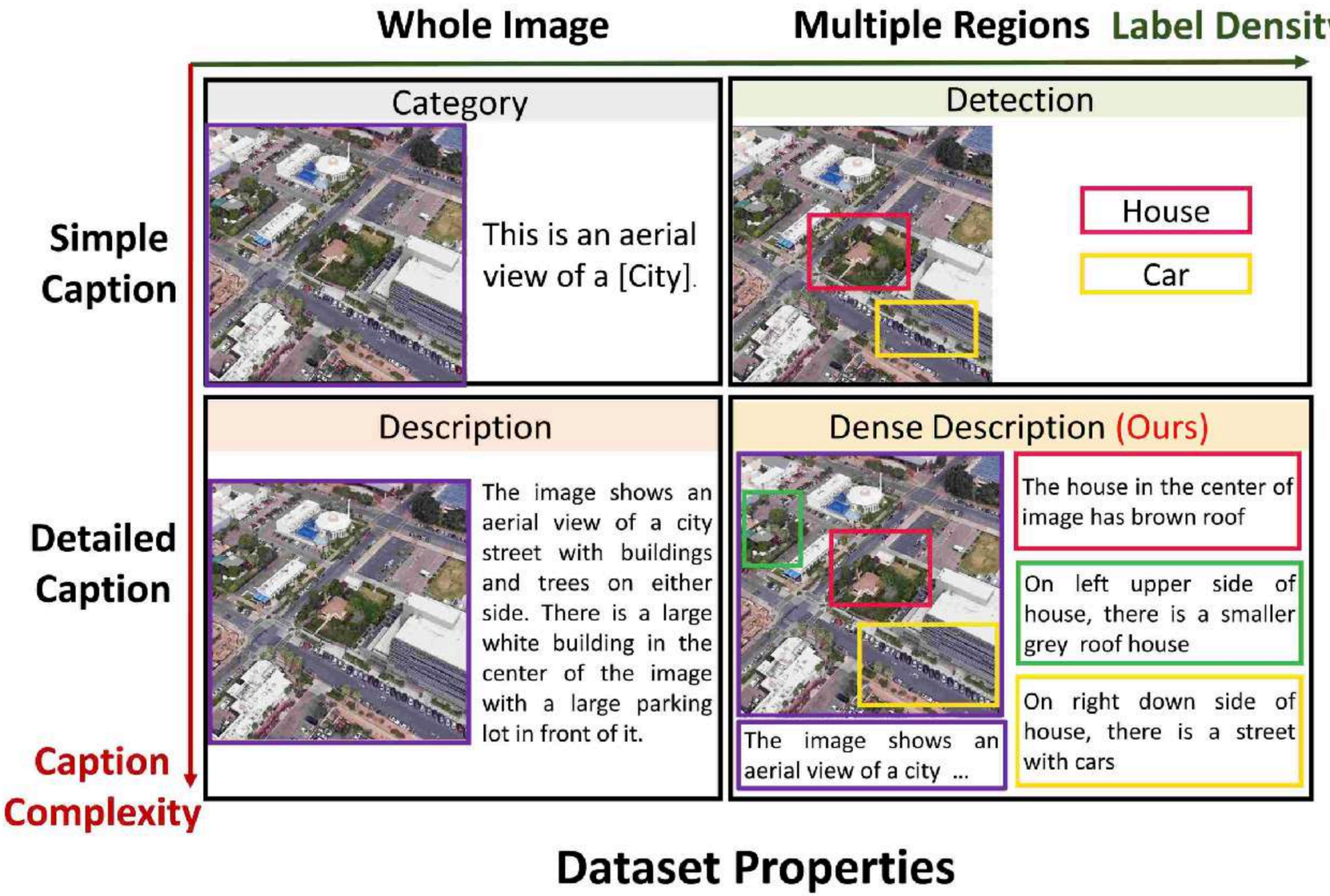
# Towards Natural Language-Guided Drones: GeoText-1652 Benchmark with Spatial Relation Matching

Meng Chu, Zhedong Zheng, Wei Ji, Tingyu Wang, Tat-Seng Chua

ECCV


- Me: Go to the pizza shop besides the school, get my pizza!
- Drone: Roger that!

# What is ideal? Caption with Spatial Relationships

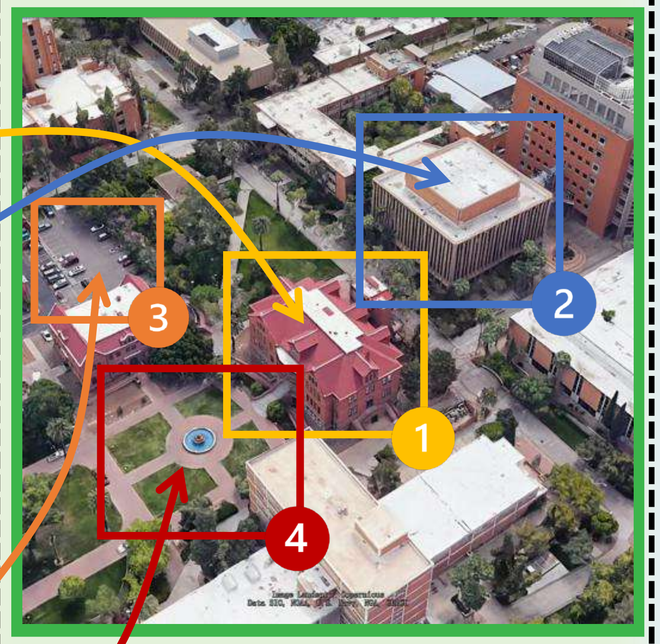


# Data Sample

**Text Query**

[1] The building has pointed red roof *in center of the image.*
[2] There is a cuboid-shaped white roof building with trees *in upper right side of central building.*
[3] There are some cars parking *in upper left side of central building.*
[4] There is a lawn with a fountain *in down left side of central building.*

**Image Retrieval**

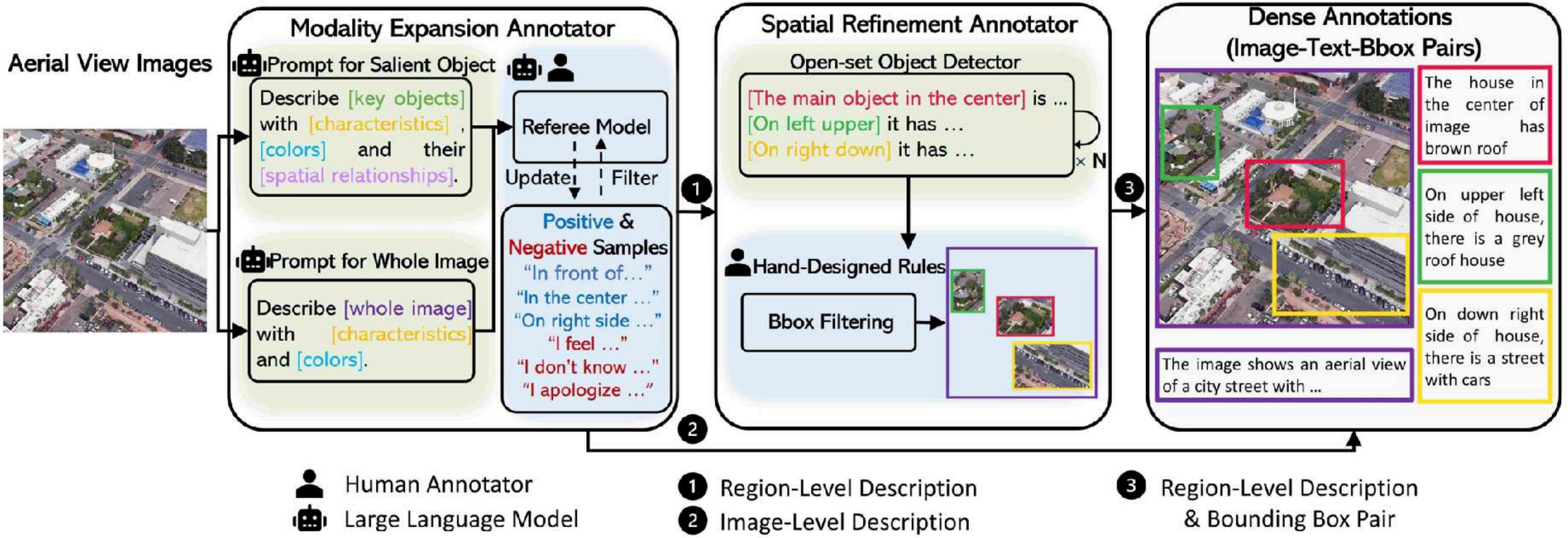



Match image features with *spatial relations*

**Geolocation Retrieval**

Find the Geolocation of Candidate: Old Main in Arizona State University, 33.4205° N, 111.9340° W

# Semi-automatic Annotation



| Split | #imgs | #global deps | #Bbox-Texts | #classes | #unis |
|---|---|---|---|---|---|
| $\text{Training}_{drone}$ | 37,854 | 113,562 | 113,367 | 701 | |
| $\text{Training}_{satellite}$ | 701 | 2,103 | 1,709 | 701 | 33 |
| $\text{Training}_{ground}$ | 11,663 | 34,989 | 14,761 | 701 | |
| $\text{Test}_{drone}$ | 51,355 | 154,065 | 140,179 | 951 | |
| $\text{Test}_{satellite}$ | 951 | 2,853 | 2,006 | 951 | 39 |
| $\text{Test}_{ground}$ | 2,921 | 8,763 | 4,023 | 793 | |

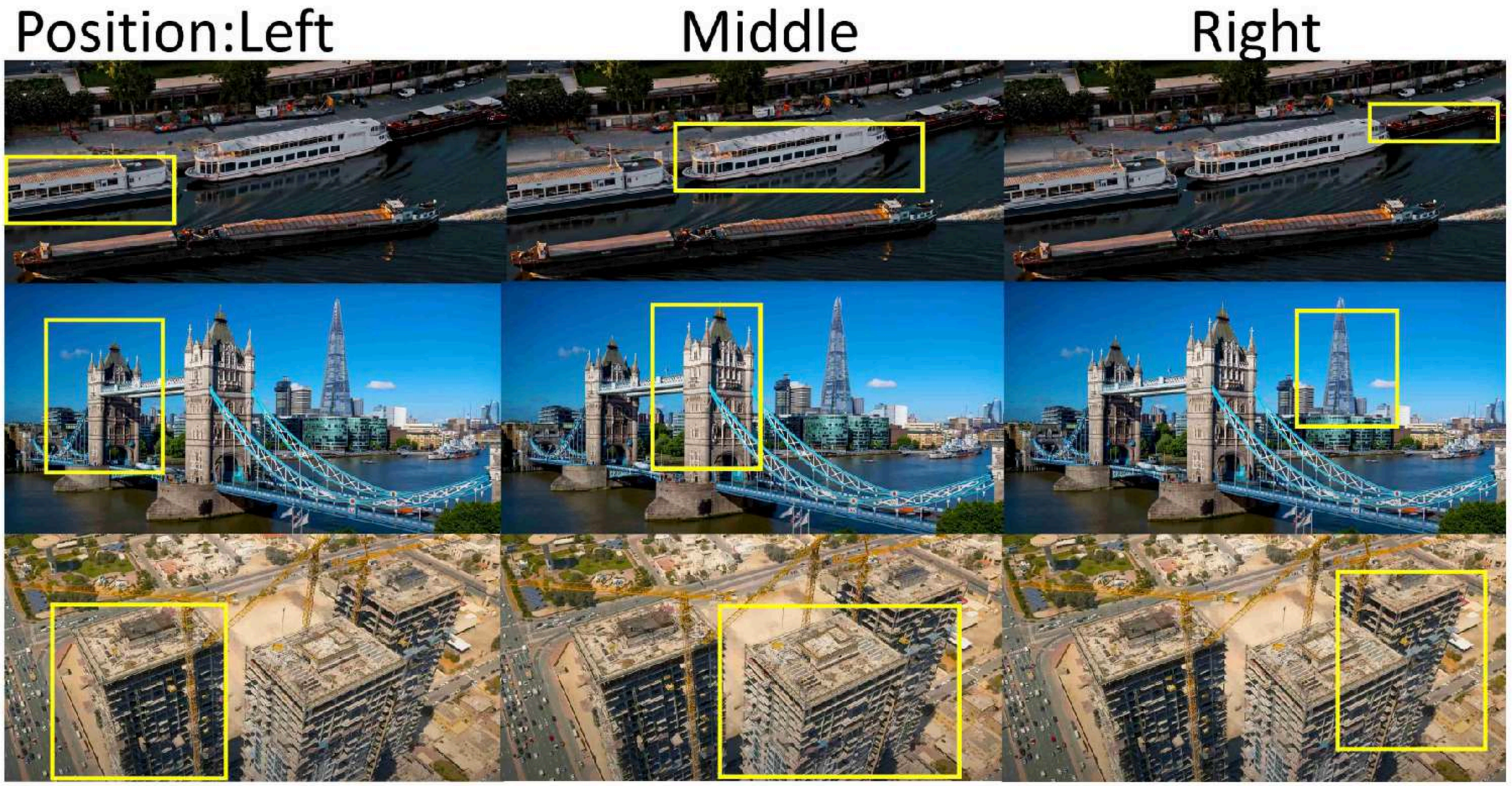
**Statics of GeoText-1652**: Training and test sets all include the image, global description, bounding box and building numbers.

| Property | CVUSA[42] | CVACT[22] | VIGOR[53] | GeoText-1652 |
|---|---|---|---|---|
| Annotation | GPS Tag | GPS Tag | GPS Tag | Sentence |
| # Bbox-Texts | N/A | N/A | N/A | 276,045 |
| Platform | G,S | G,S | G,S | G,S,D |
| Modality | Image | Image | Image | Image, Text |

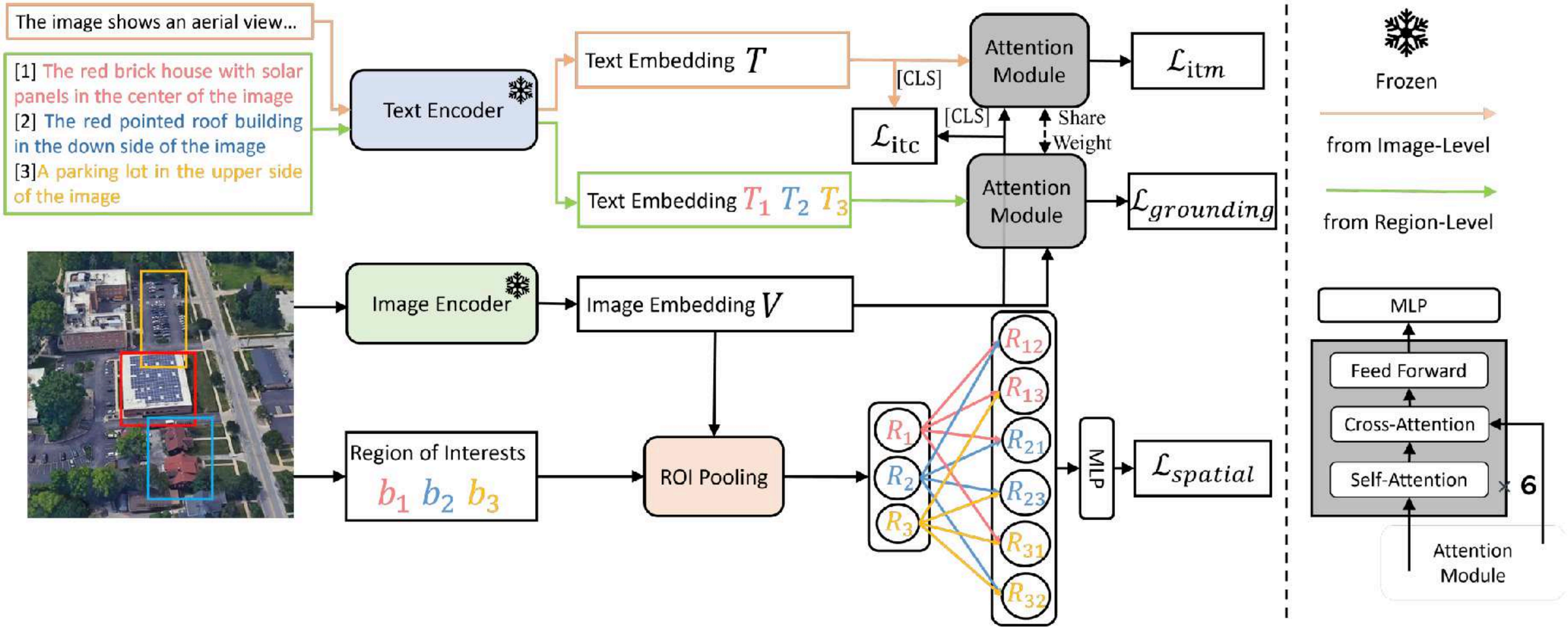
**Comparison between datasets**: G, S, and D denote ground-view, satellite-view and drone-view image, respectively.

# Why spatial-relation?  What dis-advantages?

## - Never rotation/fliplr.



# Framework Overview



# Spatial Relation Learning

$$\mathcal{L}_{\text{grounding}} = \mathbb{E}[\mathcal{L}_{\text{iou}}(\boldsymbol{b}_j, \hat{\boldsymbol{b}}_j) + ||\boldsymbol{b}_j - \hat{\boldsymbol{b}}_j||_1].$$

Layer Perceptron (MLP) to predict the 9-class spatial relationship ${\boldsymbol{p}_{\boldsymbol{r}}}^{ij}$. The spatial loss is defined as the cross-entropy loss between ${\boldsymbol{y}_{\boldsymbol{r}}}^{ij}$ and ${\hat{\boldsymbol{p}}_{\boldsymbol{r}}}^{ij}$:

$$\mathcal{L}_{\text{spatial}} = \mathbb{E}[-{\boldsymbol{y}_{\boldsymbol{r}}}^{ij} \log({\hat{\boldsymbol{p}}_{\boldsymbol{r}}}^{ij})], \tag{5}$$

# Performance

| Method | # Params | # Pretrained Images | Text Query | | | Image Query | | |
|---|---|---|---|---|---|---|---|---|
| | | | R@1 | R@5 | R@10 | R@1 | R@5 | R@10 |
| UNITER [11] | 300M | 4M | 0.9 | 2.7 | 4.2 | 2.5 | 7.4 | 11.8 |
| METER-Swin [17] | 380M | 4M | 1.3 | 3.9 | 5.8 | 2.7 | 8.0 | 12.2 |
| ALBEF [32] | 210M | 4M | 1.8 | 4.8 | 7.1 | 2.9 | 8.1 | 12.4 |
| ALBEF [32] | 210M | 14M | 1.1 | 3.5 | 5.3 | 3.0 | 9.1 | 14.2 |
| XVLM [74] | 216M | 4M | 4.3 | 9.1 | 13.2 | 4.9 | 14.2 | 21.1 |
| XVLM [74] | 216M | 16M | 4.5 | 9.9 | 13.4 | 5.0 | 14.4 | 21.4 |
| UNITER$_{finetuned}$ | 300M | 4M | 10.6 | 20.4 | 26.1 | 21.4 | 43.4 | 59.5 |
| METER-Swin$_{finetuned}$ | 380M | 4M | 11.3 | 21.5 | 27.3 | 22.7 | 46.3 | 60.7 |
| ALBEF$_{finetuned}$ | 210M | 4M | 12.3 | 22.8 | 28.6 | 22.9 | 49.5 | 62.3 |
| ALBEF$_{finetuned}$ | 210M | 14M | 12.5 | 22.8 | 28.5 | 23.2 | 49.7 | 62.4 |
| XVLM$_{finetuned}$ | 216M | 4M | 13.1 | 23.5 | 29.2 | 23.6 | 50.0 | 63.2 |
| XVLM$_{finetuned}$ | 216M | 16M | 13.2 | 23.7 | 29.6 | 25.0 | 52.3 | 65.1 |
| Ours | 217M | [illegible] | **13.6** | **24.6** | **31.2** | **26.3** | **53.7** | **66.9** |

**Table 2: Image-text bi-direction retrieval results on GeoText-1652.** Text Query: Drone Navigation (Text-to-Image Search). Image Query: Drone-view Geolocalization (Image-to-Text Search). We adopt Recall@K as the evaluation metric.

# Performance

**Table 3:** Ablation studies on: **(a)** Spatial and bbox losses. **(b)** Different training sets. **(c)** The hyper-parameter $\lambda$ selection. **(d)** Rotation angles.

**(a)**

| Method | Text Query | | | Image Query | | |
|---|---|---|---|---|---|---|
| | R@1 | R@5 | R@10 | R@1 | R@5 | R@10 |
| Baseline [74] | 13.2 | 23.7 | 29.6 | 25.0 | 52.3 | 65.1 |
| $w$ grounding loss | 13.5 | 24.4 | 30.9 | 25.9 | 53.4 | 66.3 |
| $w$ spatial loss | 13.4 | 24.0 | 30.1 | 25.3 | 52.8 | 65.6 |
| Ours | **13.6** | **24.6** | **31.2** | **26.3** | **53.7** | **66.9** |

**(b)**

| Training Set | #imgs | Text Query | | | Image Query | | |
|---|---|---|---|---|---|---|---|
| | | R@1 | R@5 | R@10 | R@1 | R@5 | R@10 |
| Drone | 37,854 | 12.9 | 23.4 | 29.1 | 25.7 | 51.5 | 64.3 |
| Satellite + Ground | 12,364 | 10.1 | 19.3 | 24.4 | 18.7 | 39.6 | 51.2 |
| Satellite + Drone + Ground | 50,218 | **13.6** | **24.6** | **31.2** | **26.3** | **53.7** | **66.9** |

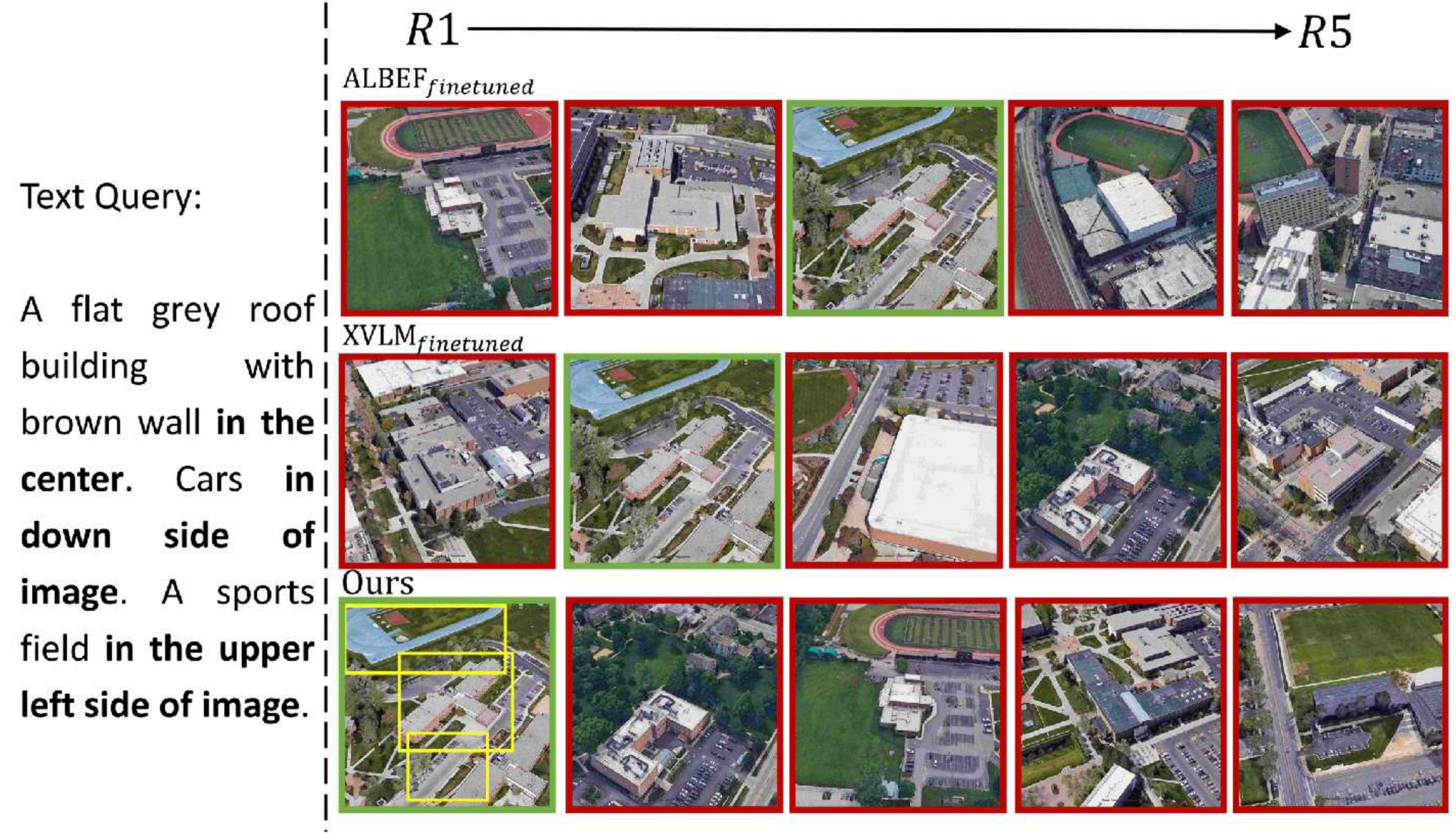
**(c)**

| $\lambda$ | Text Query | | | Image Query | | |
|---|---|---|---|---|---|---|
| | R@1 | R@5 | R@10 | R@1 | R@5 | R@10 |
| 1.00 | 10.5 | 21.6 | 27.6 | 21.8 | 47.5 | 60.8 |
| 0.50 | 11.2 | 22.3 | 29.4 | 23.2 | 51.4 | 63.5 |
| 0.10 | **13.6** | **24.6** | **31.2** | **26.3** | **53.7** | **66.9** |
| 0.05 | 12.3 | 24.1 | 30.6 | 24.6 | 52.9 | 65.2 |

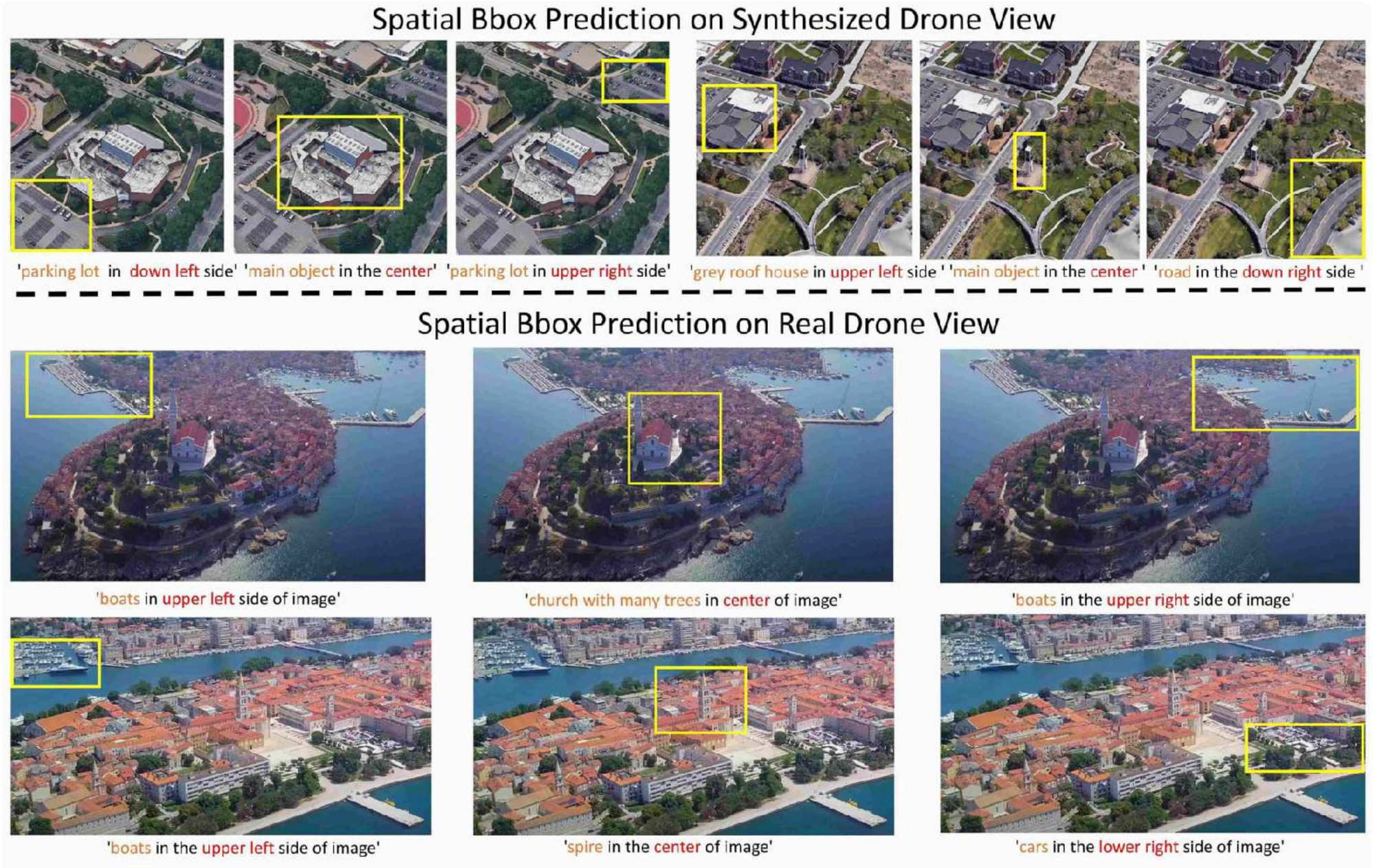
**(d)**

| Rotation Degree | Ours | | | Baseline | | |
|---|---|---|---|---|---|---|
| | R@1 | R@5 | R@10 | R@1 | R@5 | R@10 |
| 0 | **13.6** | **24.6** | **31.2** | 13.2 | 23.7 | 29.6 |
| 15 | 13.4 | 24.3 | 30.9 | 13.0 | 23.6 | 29.4 |
| 90 | 13.1 | 23.7 | 29.6 | 12.9 | 23.4 | 29.1 |
| 180 | 13.3 | 23.9 | 30.2 | 13.1 | 23.6 | 29.5 |
| 270 | 13.2 | 23.8 | 29.8 | 12.9 | 23.5 | 29.2 |

# Performance



# Scalability



澳門大學
UNIVERSIDADE DE MACAU
UNIVERSITY OF MACAU

# Thanks a lot!

## The 3rd Workshop (Multimedia UAVs: Capturing the World from a New Perspective)

**Zhedong Zheng**
University of Macau

Dataset & Code