UNIVERSITY OF ZAGREB
FACULTY OF ELECTRICAL ENGINEERING AND COMPUTING
DIGITAL IMAGE PROCESSING AND ANALYSIS

VPR VISUAL PLACE RECOGNITION

Karmela Arambašić
Kristo Palić
Nikolina Špehar
Ana Žanko
Veronika Žunar

Zagreb, June 2024

VISUAL PLACE RECOGNITION

- various domains of application
- using visual information
- role of deep learning and computer vision
- CNN models and specialized VPR models

OVERVIEW OF METHODS

ResNet

- „residual” connections
- output directly passed to the next layer

NetVLAD

- efficient aggregation of local descriptors into a powerful global image representation

DenseNet

- connecting each layer to all previous layers
- richer set of features

TransVPR

- transformer architecture
- frequent changes in visual information

RELATED WORK

- comparation of neural networks and transformers
- transformers proved to be much better
  - mechanism to process all parts of the input data simultaneously
- sequence matching
- sequential descriptors
- global descriptors

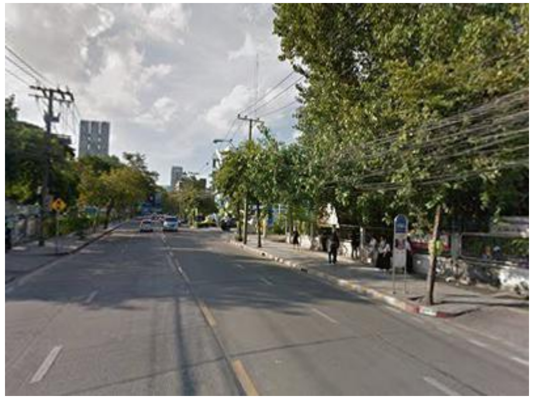
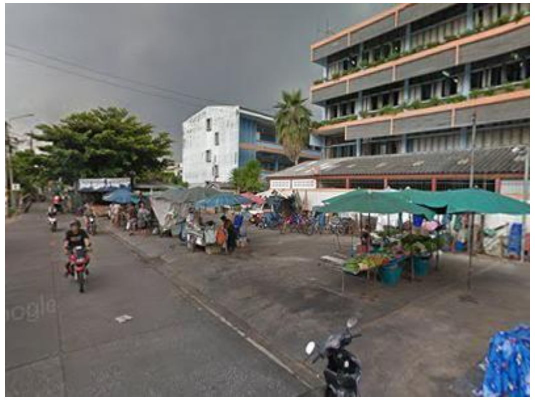
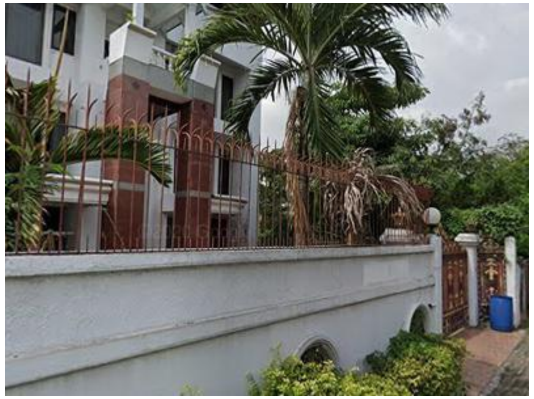
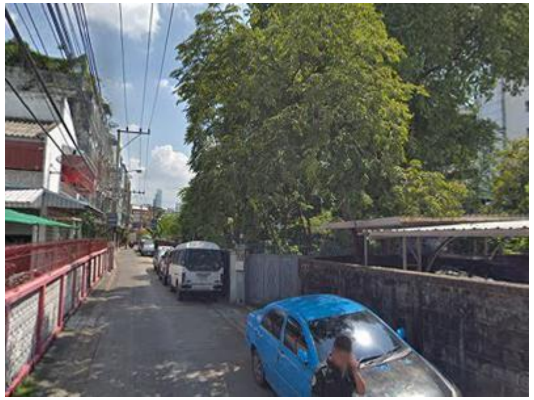
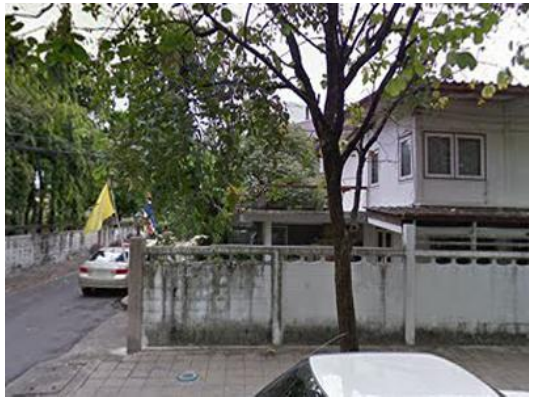
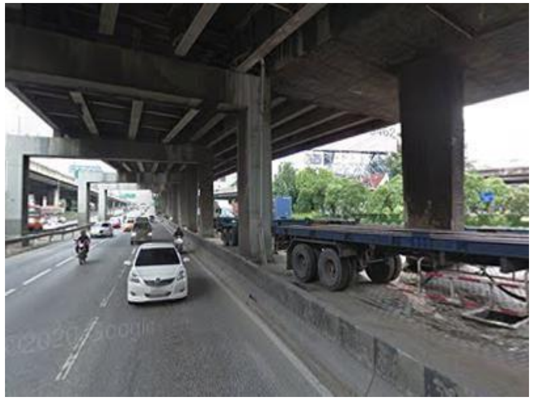
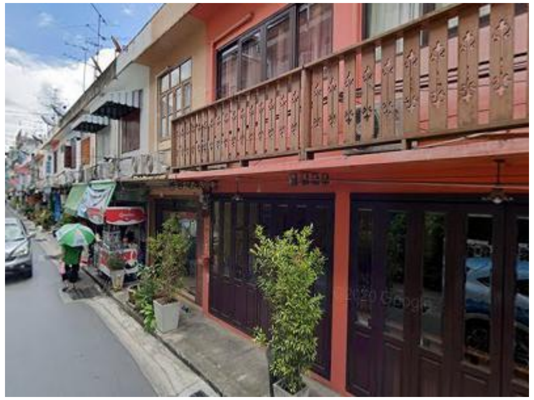
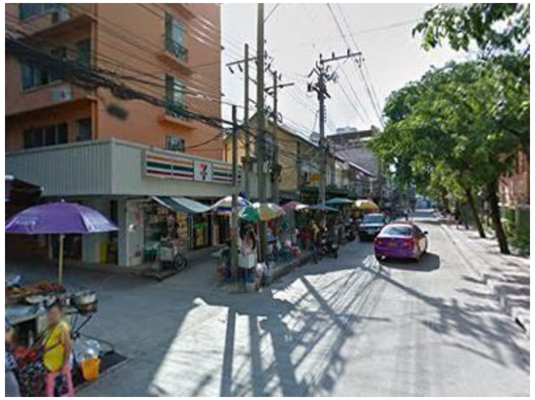
USED MATERIALS AND METHODS

1. DATASET
2. IMPLEMENTED SOLUTION
3. RESULTS

DATASET – GSV-CITIES

- *Neurocomputing 2022: GSV-Cities: Toward Appropriate Supervised Visual Place Recognition*
- 530 000 images from over 62 000 different locations
- common image dataset problems:
    - geographical coverage
    - precision of reference dana
    - perceptual diversity

GSV-CITIES – BANKOK

IMPLEMENTED SOLUTION

1. DATA PREPARATION
2. SELECTION OF OPTIMAL HYPERPARAMETERS
3. MODEL DEFINITION
4. MODEL TRAINING
5. MODEL EVALUATION

1.

DATA PREPARATION

- CustomImageDataset
- image transformations:
    - resizing to 224x224 pixels
    - random horizontal flipping
    - normalization
    - conversion to tensors

2.

SELECTION OF OPTIMAL HYPERPARAMETERS

- different learning rates
- different optimization algorithms (SGD and Adam)
- different batch sizes
- 5% of the data

3. MODEL DEFINITION

- pre-trained ResNet18 model

4. MODEL TRAINING

- cross-validation and EarlyStopping
- 50 epochs

5. MODEL EVALUATION

- overall loss and accuracy on training, validation and test sets

RESULTS

TRAINING	VALIDATION	TEST
LOSS	LOSS	LOSS
0.0308	**0.2678**	**0.2693**
ACCURACY	ACCURACY	ACCURACY
99.20%	**91.49%**	**91.53%**

CONCLUSION

- significant enhancement of VPR system capabilities
- high efficiency demonstrated on the GSV-Cities dataset
- potential for better results with further research and refinement of transformer-based models
- wide range of possible applications of VPR

VPR
VISUAL PLACE RECOGNITION

THANK YOU!
FOR YOUR ATTENTION

Karmela Arambašić
Kristo Palić
Nikolina Špehar
Ana Žanko
Veronika Žunar

UNIVERSITY OF ZAGREB
FACULTY OF ELECTRICAL ENGINEERING AND COMPUTING
DIGITAL IMAGE PROCESSING AND ANALYSIS

Zagreb, June 2024

LITERATURE

[1] R. Mereu, G. Trivigno, G. Berton, C. Masone, B. Caputo, "**Learning Sequential Descriptors for Sequence-based Visual Place Recognition**", 2022.

[2] S. Sundar Kannan, B. Min, "**PlaceFormer: Transformer-based Visual Place Recognition using Multi-Scale Patch Selection and Fusion**", 2024.

[3] A. Ali-bey, B. Chaib-draa, P. Giguère "**GSV-CITIES: Toward Appropriate Supervised Visual Place Recognition**"