Final Project
CSE 327

Dowon Kim (114001605)

Introduction

+ SNN

Introduction

Introduction

→ "Tokyo"

Motivation

"Guess where you are based on your surroundings"

Motivation


GEOGUESSR
MAP
World
ROUND
3 / 5
SCORE
6,009
ROUND 1
0:46
ROUND 2
0:32
ROUND 3
0:15
ROUND 4
ROUND 5
TOTAL
1:34
PEIXE
CAPIM GROSSO


NORTH AMERICA
EUROPE
ASIA
Atlantic Ocean
AFRICA
SOUTH AMERICA
Indian Ocean
OCEANIA
Pacific Ocean
NORTH AMERICA
Atlantic Ocean
SOUTH AMERICA
1 point
Your guess was 16,739 KM from the correct location.

Dataset

Five locations

- Seoul, Tokyo, NYC, London, Paris
- Big, populous cities.
- Well-documented coverage.

Dataset

Size/Quantity

- 50,000 images
- 32 x 32
- 3 color channels (RGB)

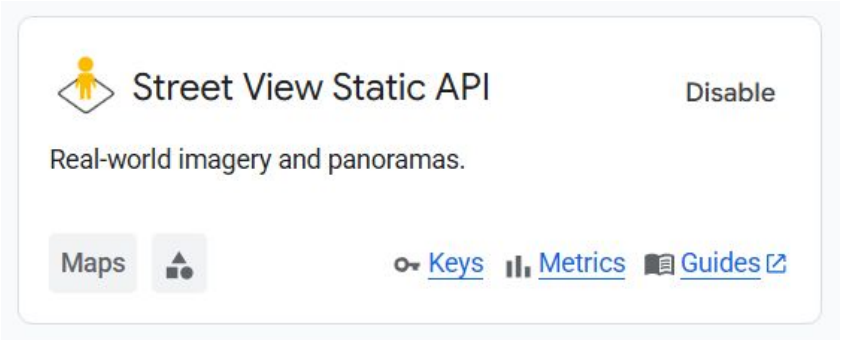
Preprocessing

- 128 x 128 → 32 x 32
- Removes huge watermarks

Dataset

Preparation

- Google Maps API
- Google streetview scraper



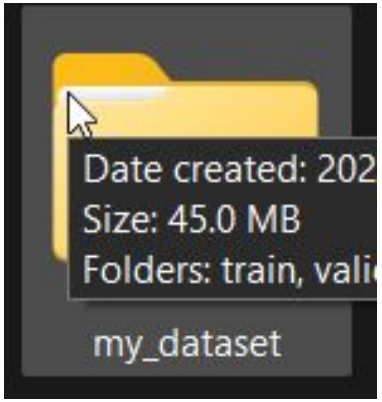
Dataset

```python
response = requests.get(signed_url)

if response.status_code == 200:
    filename = os.path.join(OUTPUT_FOLDER, f"streetview_{index}.jpg")

    # Save original image
    with open(filename, "wb") as file:
        file.write(response.content)

    # Open and resize the image to 32x32
    try:
        with Image.open(filename) as img:
            resized_img = img.resize((FINAL_SIZE, FINAL_SIZE), Image.Resampling.LANCZOS)
            resized_img.save(filename)
```

1-2 hours, 2 Google accounts

Dataset

- **my_dataset**
  - **train**
    - class0
      - images 0-7999
    - class1
      - images 0-7999
    - …
  - **validation**
    - class0
      - images 8000-9999
    - class1
      - images 8000-9999
    - …



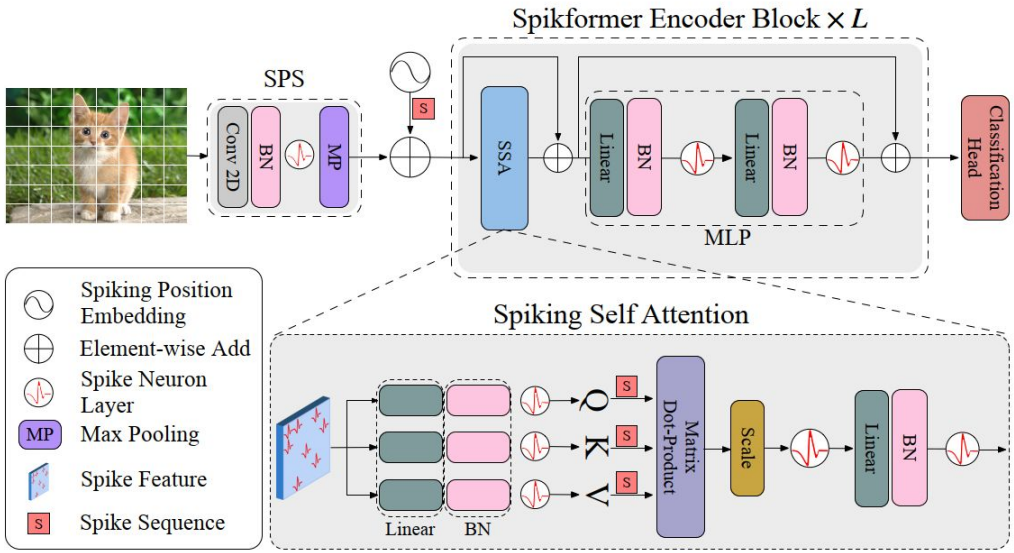
Model Architecture

Spikformer (2023)
→ Spiking neural network + Transformer

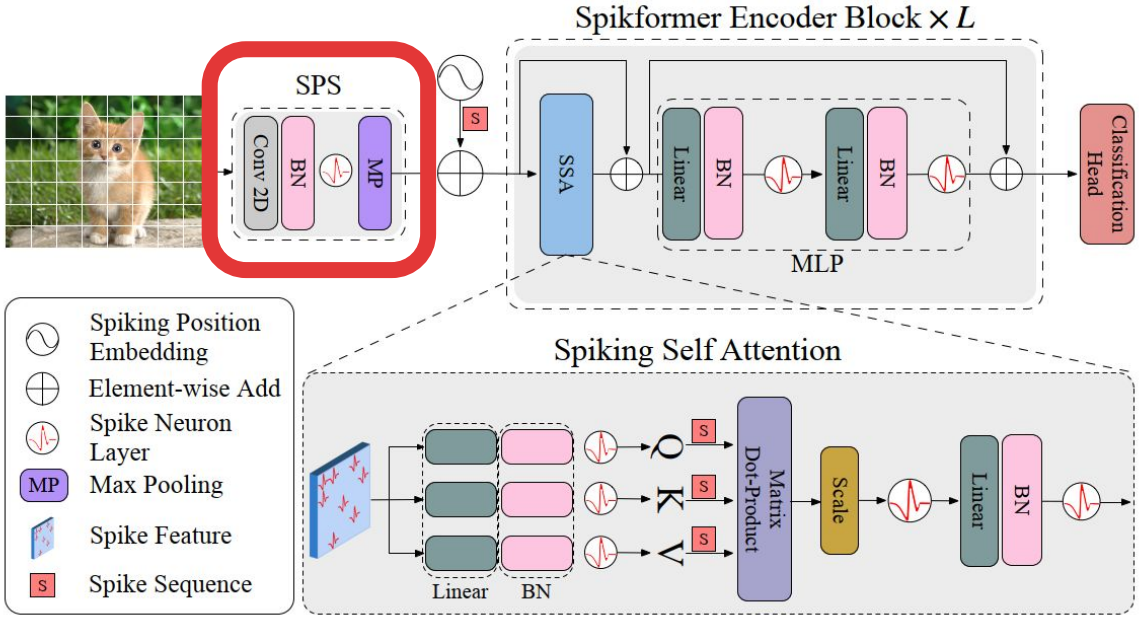
Foundation for various SNN transformer models in the future.

Model Architecture

Spikformer: When Spiking Neural Network Meets Transformer

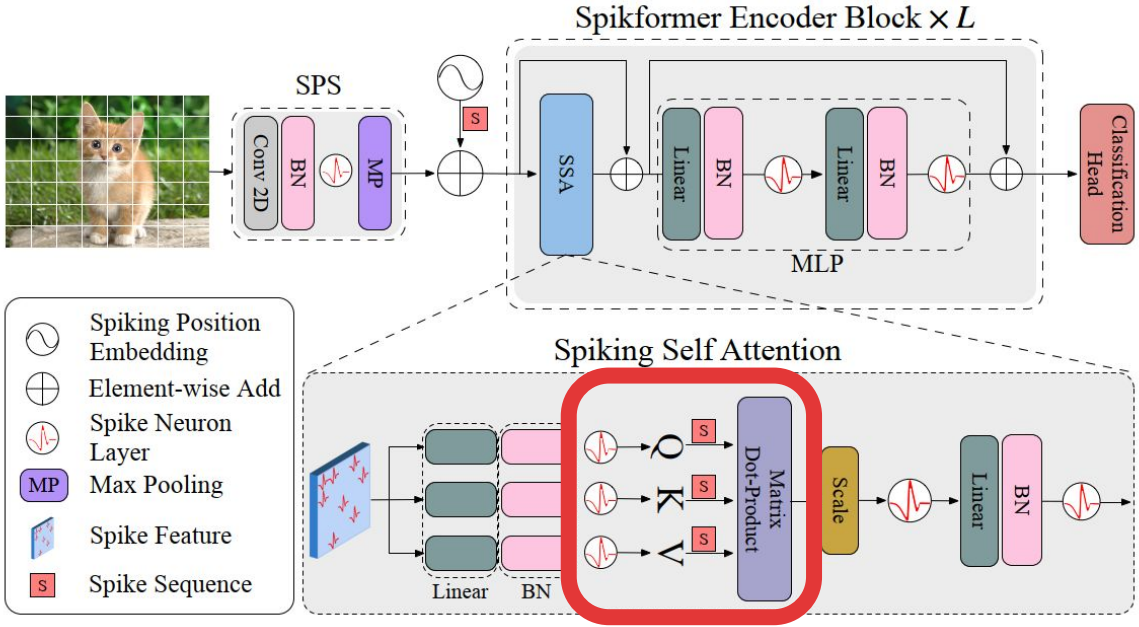


Model Architecture



SPS: Splits input into N flattened spike patches

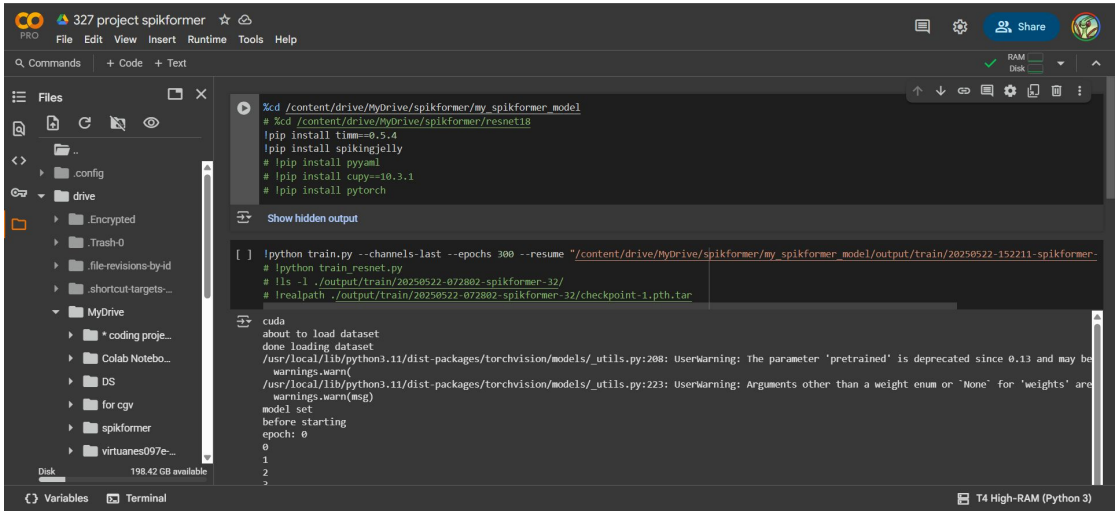
Model Architecture



SSA: Perform dot product on sparse spike matrices (no floating points)

Training Environment

Google Colab + Google Drive
T4 GPU - around 4 hours of free use



Training Environment

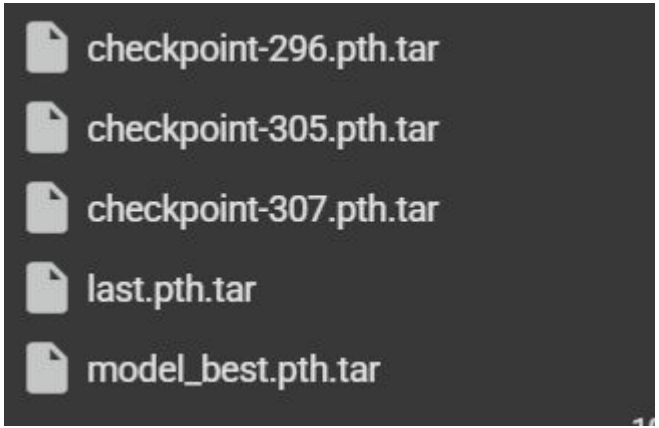
Hyperparameters

- epochs: 300
- time_step: 4
- patch_size: 4
- batch_size: 64
- lr: 5e-4
- opt: adamw

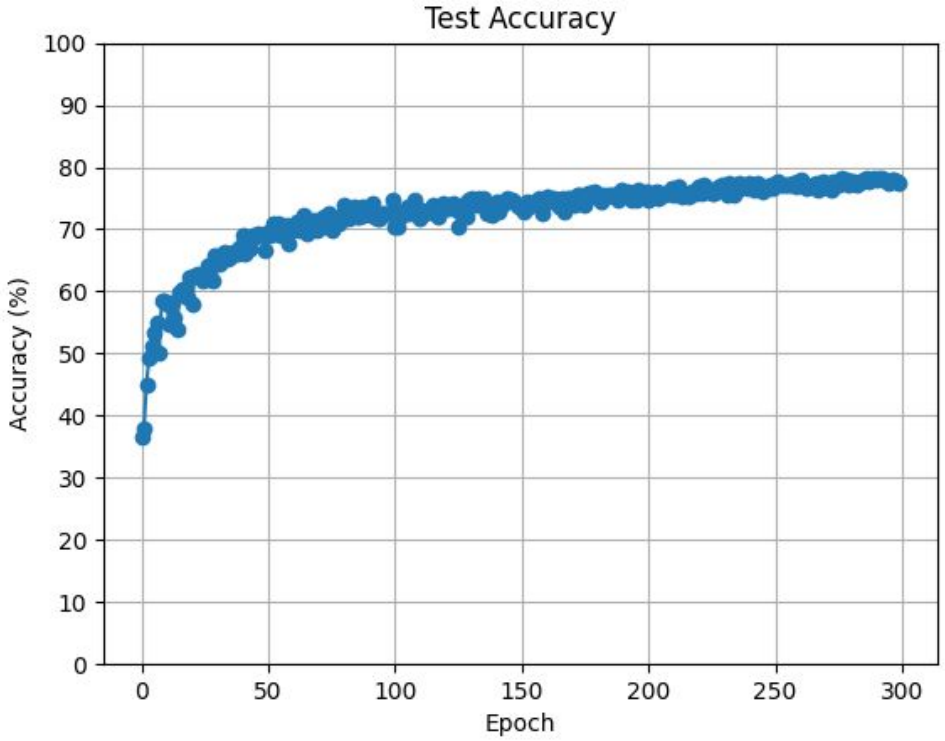
Training Procedure

Run train.py on Colab

Resume training with checkpoints



Results



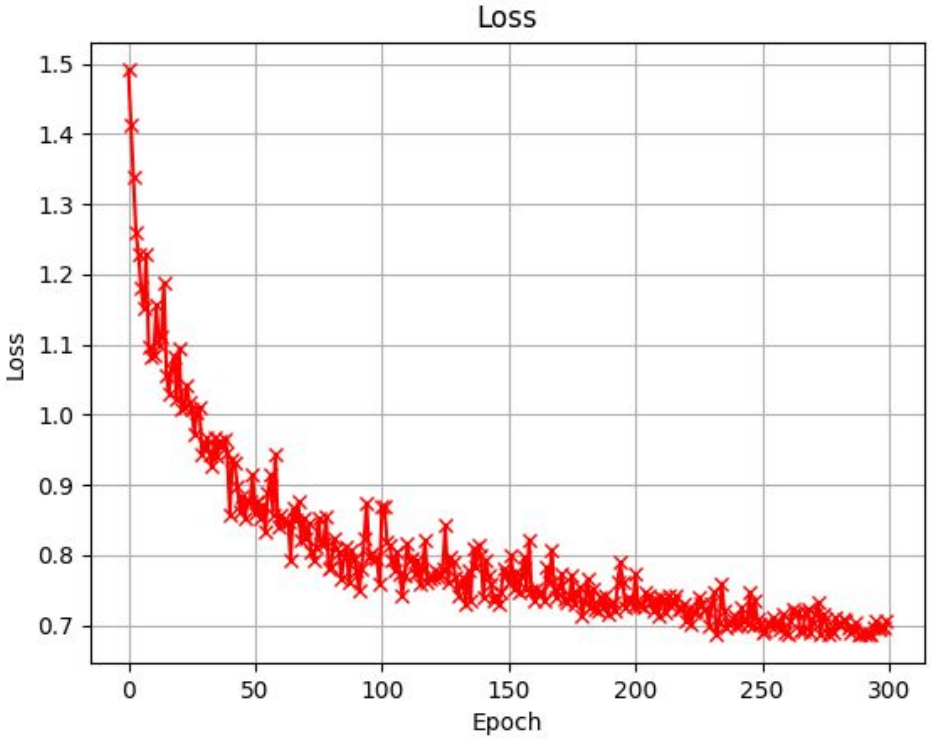
Highest accuracy:

- 78.24%

Total time:

- 15h 39m

Results



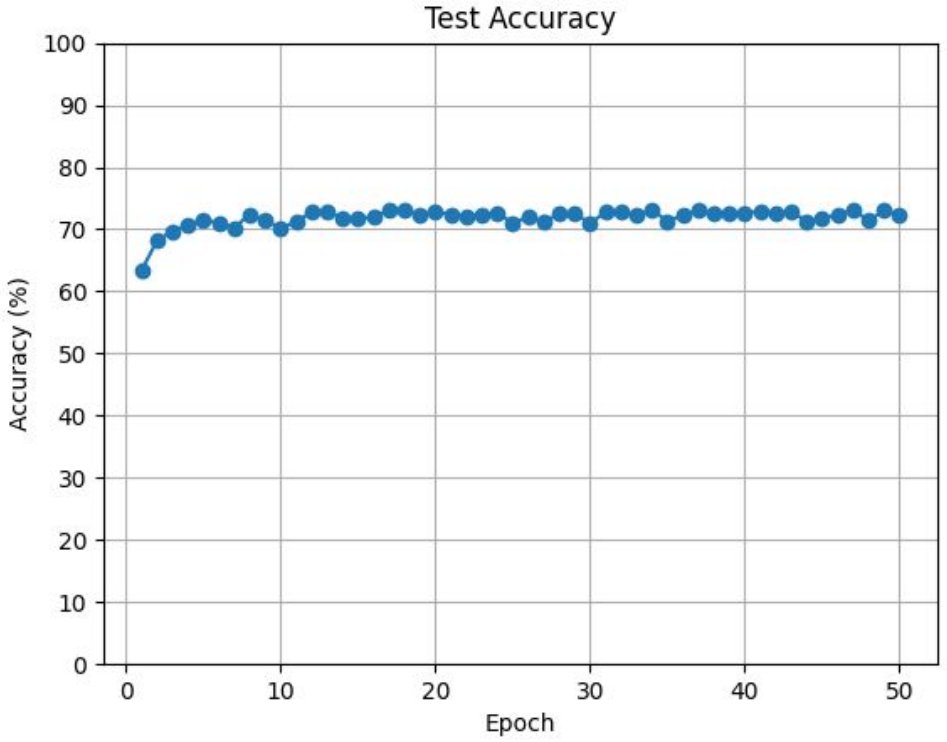
Highest accuracy:

- 78.24%

Total time:

- 15h 39m

Model Comparison - ResNet18

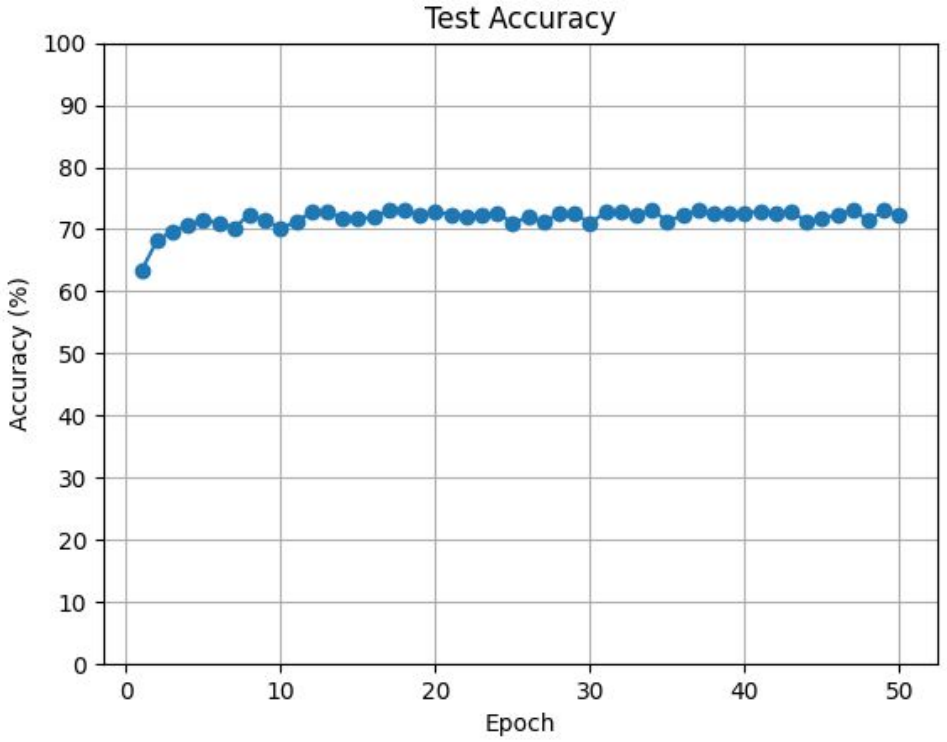


Highest accuracy:

- 73.16%
- ~4% lower

Performance stabilized very quickly

Model Comparison - ResNet18



Time per epoch

- **ResNet18**: 364.18s
- **Spikformer**: 187.92s

Me vs Spikformer

Me vs Spikformer

My guess: London
Spikformer: London

Me vs Spikformer

My guess: London
Spikformer: London

Answer: London

Me vs Spikformer

Me vs Spikformer

My guess: Seoul
Spikformer: Tokyo

Me vs Spikformer

My guess: Seoul
Spikformer: Tokyo

Answer: Seoul

Me vs Spikformer

Me: 14/25

Spikformer: 19/25

Challenges

- Training environment had more issues than expected
- Could not monitor energy output

Conclusion

- Results exceeded my expectations
- Gained insights into various aspects of training a vision model
- Felt motivated & enjoyed the process

References

Deng, B., et al. (2022). Spikformer: When Spiking Neural Network Meets Vision Transformer. arXiv: 2209.15425. https://arxiv.org/abs/2209.15425

Yao, M., Hu, J., Zhou, Z., Yuan, L., Tian, Y., Xu, B., & Li, G. (2023). Spike-driven Transformer. arXiv:2307.01694. https://arxiv.org/abs/2307.01694

Haas, L., Skreta, M., Alberti, S., & Finn, C. (2024). PIGEON: Predicting Image Geolocations. arXiv:2307.05845. https://arxiv.org/abs/2307.05845

GeoGuessr. (n.d.). GeoGuessr - Let's explore the world!. Retrieved from https://www.geoguessr.com

Thank You!

Any questions?