SLICING AIDED HYPER INFERENCe AND FINE-TUNING FOR SMALL OBJECT DETECTION

Fatih Cagatay Akyon¹,², Sinan Onur Altinuc¹,², Alptekin Temizel²
¹OBSS AI, Ankara, Turkey
²Graduate School of Informatics, Middle East Technical University, Ankara, Turkey
Problem Definition

detecting **small objects**

5-10 px

in **large images**

1000+ px
Motivation

Pretraining with large objects on low-res images

COCO pretraining

model with random weights

model with general purpose weights

small object fine-tuning

model with task-specific weights

Fine-tuning with small objects on high-res images

update architecture for small objects
Motivation

- Model with random weights
- Model with pretraining
- Model with general purpose weights
- Small object fine-tuning
- Model with task-specific weights

Fine-tune with large input size
Motivation

- Requires large GPU memory
- Low GPU utilization
Slicing Aided Fine-tuning

Pretraining dataset

Fine-tuning dataset

Augmented fine-tuning dataset

Pretrained model

Fine-tuned model

Slicing aided fine-tuning (SF)
Slicing Aided Hyper Inference (SAHI)
Experiment Setup

Datasets:
- Visdrone:
  - 10 object categories
  - 6471 training images
- xView:
  - 60 object categories
  - 846 training images

Training Framework:
- Pytroch (v1.10.0)
- MMDetection (v2.21.0)
Experiment Setup

Object Detection Models:
- FCOS: Fully Convolutional One-Stage Object Detection
  - Anchor box free, Eliminates anchor-box related hyperparameters
  - Only requires NMS as post-processing

- VarifocalNet: An IoU-aware Dense Object Detector
  - Learns to predict the IoU-aware classification score which mixes the object presence confidence and localization accuracy together as the detection score for a bounding box.

- TOOD: Task-aligned One-stage Object Detection
  - Explicitly aligns the two tasks in a learning-based manner.

Notations:
- FI: Full-Image inference
- SAHI: Slicing aided inference
- PO: Patch Overlap
- SF: Slicing aided fine-tuning
Evaluation Results: Visdrone Dataset
Evaluation Results: Visdrone Dataset

| Setup                  | AP$_{50}$ | AP$_{50s}$ | AP$_{50m}$ | AP$_{50l}$ |
|------------------------|-----------|------------|------------|------------|
| FCOS+FI                | 25.8      | 14.2       | 39.6       | 45.1       |
| FCOS+SAHI+PO           | 29.0      | 18.9       | 41.5       | 46.4       |
| FCOS+SAHI+FI+PO        | 31.0      | 19.8       | 44.6       | 49.0       |
| FCOS+SF+SAHI+PO        | 38.1      | 25.7       | 54.8       | 56.9       |
| FCOS+SF+SAHI+FI+PO     | **38.5**  | **25.9**   | **55.4**   | **59.8**   |
| VNet+FI                | 28.8      | 16.8       | 44.0       | 47.5       |
| VNet+SAHI+PO           | 32.0      | 21.4       | 45.8       | 45.5       |
| VNet+SAHI+FI+PO        | 33.9      | 22.4       | 49.1       | 49.4       |
| VNet+SF+SAHI+PO        | 41.9      | **29.7**   | 58.8       | 60.6       |
| VNet+SF+SAHI+FI+PO     | **42.2**  | **29.6**   | **59.2**   | **63.3**   |
| TOOD+FI                | 29.4      | 18.1       | 44.1       | 50.0       |
| TOOD+SAHI              | 31.9      | 22.6       | 44.0       | 45.2       |
| TOOD+SAHI+PO           | 32.5      | 22.8       | 45.2       | 43.6       |
| TOOD+SAHI+FI           | 34.6      | 23.8       | 48.5       | 53.1       |
| TOOD+SAHI+FI+PO        | 34.7      | 23.8       | 48.9       | 50.3       |
| TOOD+SF+FI             | 36.8      | 24.4       | 53.8       | **66.4**   |
| TOOD+SF+SAHI           | 42.5      | 31.6       | 58.0       | 61.1       |
| TOOD+SF+SAHI+PO        | 43.1      | **31.7**   | 59.0       | 60.2       |
| TOOD+SF+SAHI+FI        | 43.4      | **31.7**   | 59.6       | 65.6       |
| TOOD+SF+SAHI+FI+PO     | **43.5**  | **31.7**   | **59.8**   | **65.4**   |

- SAHI increases object detection AP by up to 6.8%.
- With SF, object detection AP increases up to 14.5% AP.
- Applying 25% overlap between slices during inference, increases small/medium object AP and overall AP.
## Evaluation Results: xView Dataset

| Setup                     | $AP_{50}$ | $AP_{50s}$ | $AP_{50m}$ | $AP_{50l}$ |
|---------------------------|-----------|------------|------------|------------|
| FCOS+FI                   | 2.20      | 0.10       | 1.80       | 7.30       |
| FCOS+SF+SAHI              | 15.8      | 11.9       | 18.4       | 11.0       |
| FCOS+SF+SAHI+PO           | **17.1**  | **12.2**   | **20.2**   | 12.8       |
| FCOS+SF+SAHI+FI           | 15.7      | 11.9       | 18.4       | 14.3       |
| FCOS+SF+SAHI+FI+PO        | **17.0**  | **12.2**   | **20.2**   | **15.8**   |
| VFNet+FI                  | 2.10      | 0.50       | 1.80       | 6.80       |
| VFNet+SF+SAHI             | 16.0      | 11.9       | 17.6       | 13.1       |
| VFNet+SF+SAHI+PO          | **17.7**  | **13.7**   | **19.7**   | 15.4       |
| VFNet+SF+SAHI+FI          | 15.8      | 11.9       | 17.5       | 15.2       |
| VFNet+SF+SAHI+FI+PO       | **17.5**  | **13.7**   | **19.6**   | **17.6**   |
| TOOD+FI                   | 2.10      | 0.10       | 2.00       | 5.20       |
| TOOD+SF+SAHI              | 19.4      | 14.6       | 22.5       | 14.2       |
| TOOD+SF+SAHI+PO           | **20.6**  | **14.9**   | **23.6**   | 17.0       |
| TOOD+SF+SAHI+FI           | 19.2      | 14.6       | 22.3       | 14.7       |
| TOOD+SF+SAHI+FI+PO        | **20.4**  | **14.9**   | **23.5**   | **17.6**   |

- SAHI+FI yielded up to 3.3% increase in large object AP compared to only SAHI.
- 25% overlap between slices increase the detection AP by up to 1.7%.
Future work

- Other postprocessing techniques
- Slicing aided small instance segmentation
- Comparison with more models
- Slicing aided video object detection
F.C. Akyon, S.O. Altinuc, A. Temizel, “Slicing Aided Hyper Inference and Fine-tuning for Small Object Detection”, IEEE International Conference on Image Processing (ICIP), Oct. 2022.
Active Learning Based Synthetic Sample Selection for Endoscopic Image Classification

Alperen İnci, Ümit Mert Çağlar, Gökem Polat, Oğuz Hanoğlu, Alptekin Temizel
Graduate School of Informatics, Middle East Technical University, Ankara, Turkey
Motivation and Problem Definition

- Ulcerative Colitis is a chronic inflammatory bowel disease.
- Assessment of the severity of the disease is crucial for physicians to administer appropriate treatment for UC disease.
Data Labelling Process

UC Mayo Annotator

Progress: 76.3%  Total images to annotate: 1468

Current image: G000774803.bmp

Your annotation:
- [ ] Değerlendirmeye uygun değil
- [ ] Mayo 0
- [ ] Mayo 1
- [ ] Mayo 2
- [ ] Mayo 3

Annotate

Show annotations

OA_IE: Mayo 3

YOA: Etiketlenmemiş!
Data Labelling Process

- 572 Patients
  - 1043 Colonoscopies
  - 19537 Images

  Annotation
  - Labeled according to EMS (0-3)
  - Not suitable to make an assessment (due to debris, artifacts vs.)
  - 8060 Images
  - Differently labeled by all three reviewers
  - 201 Images
  - 8261 images were removed from the dataset

- 564 Patients
  - 11276 Images

  Model Development (~85%)
  - 479 Patients
  - 9590 Images

  10-fold Cross-Validation
  - Training Set (~76.5%)
    - ~431 Patients
    - ~8831 Images

  - Validation Set (~8.5%)
    - ~48 Patients
    - ~959 Images

  Test Set (~15%)
  - 85 Patients
  - 1686 Images

  Trained DNN Model
  - Inference on Test Set
  - Calculate Performance Metrics
Data Labelling by Subject Matter Experts

|                                | Reviewer-1 | Reviewer-2 |
|--------------------------------|------------|------------|
| Total images to evaluate       | 19537      |            |
| Not suitable to assign a Mayo score | 7621      | 9207       |
| Mayo score is assigned         | 11916      | 10330      |
| Mayo-0                         | 7398       | 4503       |
| Mayo-1                         | 2473       | 3796       |
| Mayo-2                         | 1190       | 1014       |
| Mayo-3                         | 855        | 1017       |
Data Labelling by Subject Matter Experts

|                              | Reviewer-3 | From Reviewer 1&2 | Total |
|------------------------------|------------|-------------------|-------|
| Total images to evaluate     | 7652       | -                 | -     |
| Not suitable to assign a Mayo score | 1895       | -                 | -     |
| Mayo score is assigned       | 5757       | -                 | -     |
| All reviewers annotate differently | 201        | -                 | -     |
| To join the final dataset    |            |                   |       |
| (agreement by two reviewers) |            |                   |       |
| Mayo-0                       | 2633       | 3472              | 6105  |
| Mayo-1                       | 1842       | 1210              | 3052  |
| Mayo-2                       | 784        | 470               | 1254  |
| Mayo-3                       | 297        | 568               | 865   |
Data Labelling by Subject Matter Experts

Histogram of number of images per patient after annotation

Number of images per Mayo subscore:

- Mayo 0: 54.14%
- Mayo 1: 27.07%
- Mayo 2: 11.12%
- Mayo 3: 7.67%
**patient_based_classified_images:** Images of each patient are separated according to Mayo classes. If a train-val-test splitting is to be made according to the ratios desired by the user, this folder should be used.

**train_and_validation_sets:** Train and validation sets used in the paper. Using the scripts in dataset's GitHub repository, same 10-fold can be generated for replicating the results.

**test_set:** Test set used for performance measurement in the research paper. For a fair performance comparisons, this should be used to report performances.
Research Questions

• When there are limited number of labelled images, can we improve model performance by generating and adding synthetic samples?
• How can we best select the synthetic samples that would be the most useful in training?

Example synthetic colonoscopy images

Mayo 0- healthy
Mayo 1-mild disease
Mayo 2-moderate disease
Mayo 3-severe disease
Method: GAN Model Training

StyleGAN2-ADA-PyTorch

• Resolution 256x256
• Training length 5M images (initially 25M)
• Best model save at 200k images
• r1 Gamma=2 (best FID among 1,2,4,8)
• All augmentations
• ADA target 0.6
• Class Conditional GANs
• Class Specific GANs
GAN Model Training

**Class Conditional GANs**
- Employs class information
- One GAN for all classes
- Better FID on original dataset (imbalanced)
- No transfer learning, trained from scratch

**Class Specific GANs**
- A separate GAN for each class
- Worse FID on original dataset
- Can apply transfer learning (FFHQ)

| Training Method | Class-Conditional | Class-Specific GAN |
|-----------------|-------------------|--------------------|
| Training set    | Mayo 0 | Mayo 1 | Mayo 2 | Mayo 3 |
| Subset 50       | 154.8  | 129.7  | 110.7  | 100.6  | 115.5  |
| Subset 100      | 128.8  | 117    | 110.7  | 102.1  | 119    |
| Subset 150      | 111.9  | 98.2   | 86.6   | 90.8   | 104    |
| Subset 200      | 94.6   | 88.7   | 77.2   | 81.6   | 92.1   |
| Subset 250      | 96.5   | 78     | 66.9   | 74     | 79.9   |
| Subset 300      | 32.4   | 70.7   | 63.4   | 66.6   | 74.6   |
| Subset 350      | 29.5   | 67.2   | 58.8   | 60.8   | 66.9   |
| Subset 400      | 23.7   | 61.8   | 53.5   | 58     | 64.8   |
| Subset 450      | 24.5   | 57.2   | 49.5   | 50.8   | 59.2   |
| Subset 500      | 18.9   | 54.1   | 50.5   | 51.1   | 55.4   |
| Original        | 8.5    | 15.1   | 21.5   | 35.8   | 53.3   |
Collective Dataset Creation

The truncation value controls the variance of generated samples.
- Truncation 0.5 - samples are mostly around distribution center
- Truncation 2.0 - samples are too diverse/unrealistic
- Truncation 1.2 - trade-off between 2.0 and 0.5.
Results: Class-Specific GAN
System Architecture

- Original images
- Synthetic images (180K samples)
- Synthetic set
- Generation Style GAN2 - ADA
- Uncertainty Sampling (Entropy, Margin)
- Diversity Sampling (Coreset)
- Active Learning Sampler
- Neural Network
- Training and Inference

Diversity Sampling (embedding space distance based)
System Architecture

- **Synthetic Images**
  - (180K samples)
  - Generated by Style GAN2 - ADA

- **Original Images**

- **Neural Network Training**
Diversity Sampling of Synthetic Images

![Diagram showing the process of diversity sampling from a synthetic image set]
Active Learning Based Sampling of Synthetic Images

- Entropy
- Coreset
- Margin
- Weighted Margin
Active Learning Based Sampling of Synthetic Images

- **Entropy**
  - Higher entropy indicates higher uncertainty - model is not confident about classification of the sample.

- **Coreset**
  - Aims to extract a diverse set of points with the maximum distance from others to represent the whole dataset.

- **Margin**
- **Weighted Margin**
Active Learning Based

- Entropy
- Coreset

Uncertainty-based active learning strategies frequently select similar samples since the trained model is likely to struggle to make decisions on almost identical samples. Therefore, uncertainty-based selection methods are prone to suffer from the overlapping problem.

- Margin
  - computes the difference between the top two class probabilities

- Weighted Margin
  - computes the uncertainty score by taking the power of Margin score with class distance
Results (50 Real Images Per Class)

Baseline QWK: 68.0
Results (50 Real Images Per Class)

Baseline F1: 54.3, Naïve Method F1: 55.8
Results (100 Real Images Per Class)

Baseline QWK: 74.6
Results (100 Real Images Per Class)

Baseline F1: 59.5, Naïve Method F1: 61.5
Conclusion

- Performance improvements can be achieved by using active learning methods.
- Comparative evaluations against random sample selection has to be done as it may outperform more sophisticated selection methods.
- Weighted Margin is the best approach according to the experimental results.