HITACHI
Inspire the Next

Hitachi at TRECVID DSDI 2020

Soichiro Okazaki, Quan Kong, Martin Klinkigt, Tomoaki Yoshinaga,
Lumada Data Science Lab., Hitachi, Ltd.

© Hitachi, Ltd. 2020. All rights reserved.

About DSDI

HITACHI
Inspire the Next

- DSDI (***D**isaster **S**cene **D**escription and **I**ndexing*) is one of the competition tasks in TRECVID 2020

[Train aerial-images]

[Unseen test short-clips]

After trained these aerial-images, create top-1000 test shot-clips ranking list for given 32 features.

- Given Dataset and Metadata information
  - **Labeled images (about 40K images / Each image is labeled by multiple AMT workers.)**
  - Unlabeled images (about 598K images)
  - Metadata information (e.g. image-size, filesize, latitude, longitude, altitude, timestamp)
  - Machine-generated labels created by ImageNet/Place365 pre-trained model

© Hitachi, Ltd. 2020. All rights reserved.

Problem Setting & Task / Training Data Examples

HITACHI
Inspire the Next

- Setting: **40672 labeled training images, 37 class (32 named class + 5 none class)**

  Task: Rank top-1000 shot clips from 1825 test shot clips for each **32 named class**

1 'damage:misc'
2 'damage:flood/water'
3 'damage:landslide'
4 'damage:washout'
5 'damage:rubble'
6 'damage:smoke/fire'
7 'environment:dirt'
8 'environment:grass'
9 'environment:lava'
10 'environment:rock'
11 'environment:sand'
12 'environment:shrubs'
13 'environment:snow/ice'
14 'environment:trees'
15 'infrastructure:bridge'
16 'infrastructure:building'
17 'infrastructure:dam-levee'
18 'infrastructure:pipe'
19 'infrastructure:utility-line'
20 'infrastructure:railway'
21 'infrastructure:communications-tower'
22 'infrastructure:water-tower'
23 'vehicle:aircraft'
24 'vehicle:boat'
25 'vehicle:car'
26 'vehicle:truck'
27 'water:flooding'
28 'water:lake'
29 'water:ocean'
30 'water:puddle'
31 'water:river'
32 'infrastructure:road'

Training Data Examples with aggregated annotators' label information

"DSC_0157_7334e9b5-d85a-405f-acfb-b058457cacb3.jpg": [
"['damage:flood/water']" / Annotator index : 389,
"['damage:flood/water']" / Annotator index : 397,
"['damage:smoke/fire']" / Annotator index : 420,
"['damage:flood/water']" / Annotator index : 432,
"['damage:flood/water']" / Annotator index : 440,
"['damage:none']" / Annotator index : 533
],

"A0013Vb0_7dfa0da4-2a81-41dd-9203-f22b9532b038.jpg": [
"['damage:rubble']" / Annotator index : 4,
"['infrastructure:building', 'infrastructure:road', 'infrastructure:utility-line']" / 5,
"['water:none']" / Annotator index : 5,
"['vehicle:car', 'vehicle:truck']" / Annotator index : 15,
"['environment:grass', 'environment:shrubs', 'environment:trees']" / 16,
"['water:none']" / Annotator index : 26,
"['vehicle:car', 'vehicle:truck']" / Annotator index : 168,

Idea behind our solution

HITACHI
Inspire the Next

- Task: Rank top-1000 shot clips from 1825 test shot clips for each 32 named class
- How to solve this task?

→Solve "Multi-label multi-class classification" task for the 32 features, and sort the predicted confidences descending to create ranking list for test shot-clips.

Ranking list	1	2	…	1000
(confidence)	0.957	0.942		0.145
flood/water			…	
(confidence)	0.487	0.479		0.077
boat			…	

© Hitachi, Ltd. 2020. All rights reserved.

Exploratory Data Analysis (Dataset Characteristics)

HITACHI
Inspire the Next

1. Noisy-annotated images

[Train sample]

[Annotation]

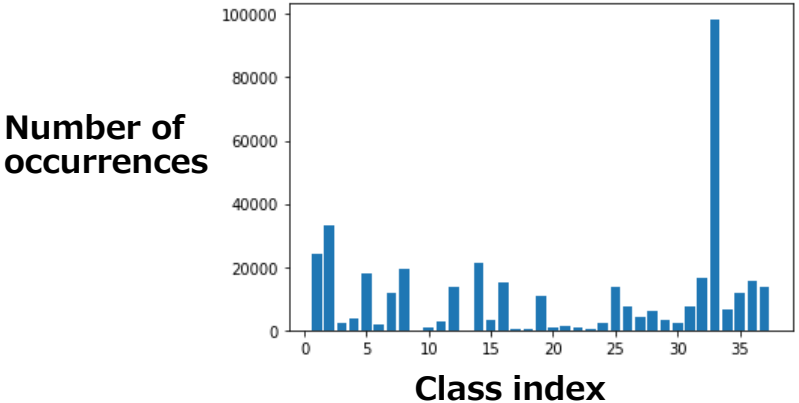
"DSC_0157_7334e9b5-d85a-405f-acfb-b058457cacb3.jpg": [
"['damage:flood/water']" / Annotator index : 389,
"['damage:flood/water']" / Annotator index : 397,
"['damage:smoke/fire']" / Annotator index : 420, ← **incorrect label**
"['damage:flood/water']" / Annotator index : 432,
"['damage:flood/water']" / Annotator index : 440,
"['damage:none']" / Annotator index : 533 ← **incorrect label**
],

[Ground Truth（※Not Provided）]

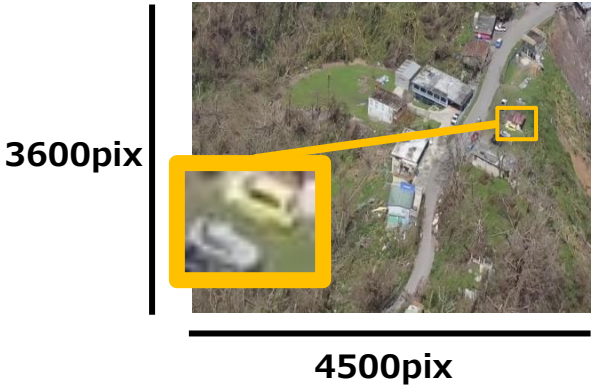
Flood/Water, Trees, Ocean, River

Missing label:
Trees, Ocean, River

2. Highly imbalanced data



3. High-Resolution images with tiny objects

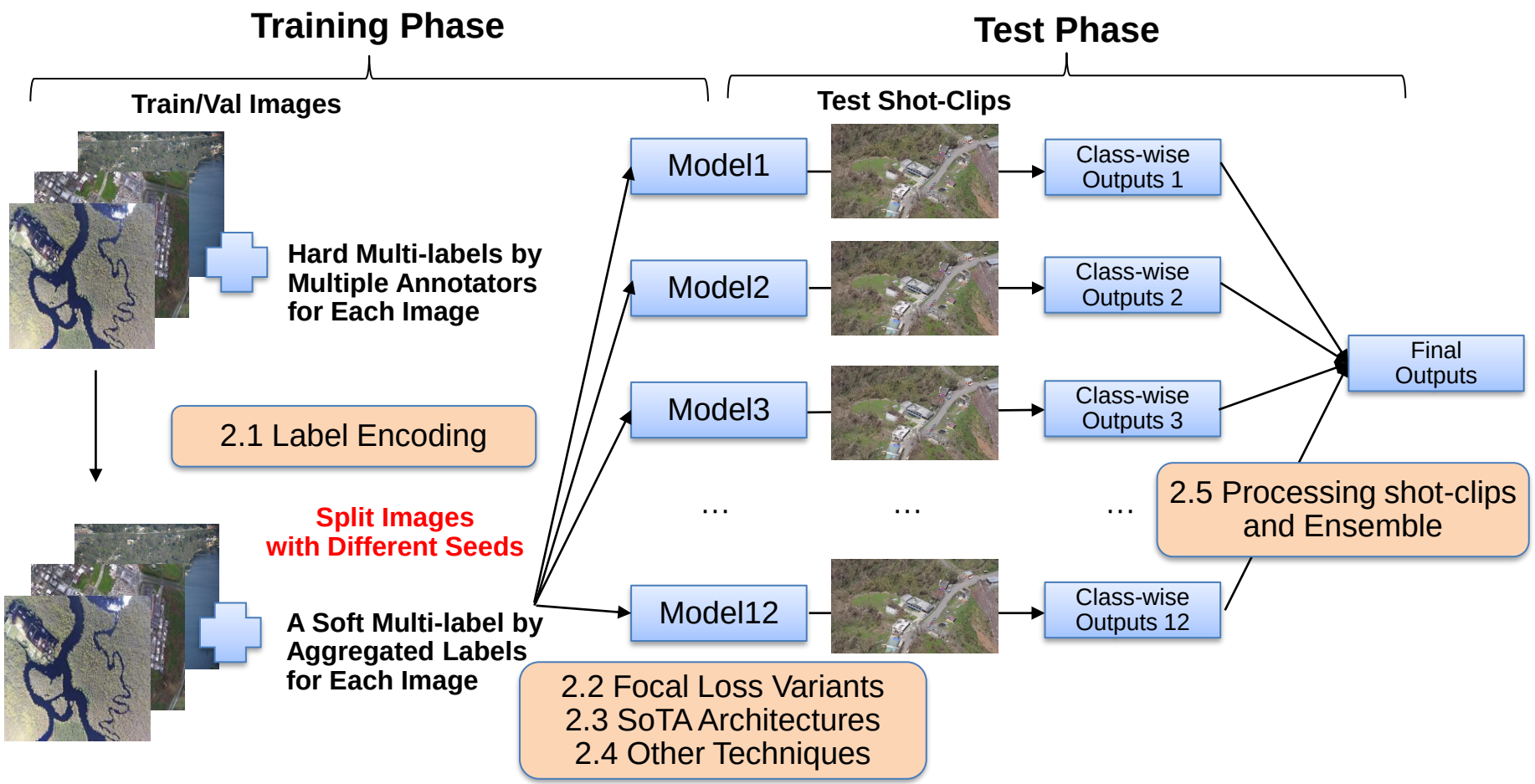


[Ground Truth]

- ✓ Rubble
- ✓ Grass
- ✓ Car
- ✓ Building
- ✓ Trees

© Hitachi, Ltd. 2020. All rights reserved.

Approach Overview



Approach Overview

HITACHI
Inspire the Next

- Our solution is based on **the ensemble of 12 models** **※A01-C12 (VAS4): mAP-0.360**
- 10 models + NICT AutoML, VAS2 achieved mAP-0.383. D-13 is for domain adaptation (but failed…)

Index	Architecture	Focal	Class-Balanced	Reduced	Other Techniques	LRAP
A01	RegNet-6.4F	$\gamma : 2.0$	-	-	-	0.846
B02	**HRNet-W18**	$\gamma : 2.0$	-	-	**Input resolution: 448*448 (default: 360*360)**	**0.854**
C03	ResNeSt-50d	$\gamma : 2.0$	$\beta : 0.999$	-	-	0.844
C04	ResNeSt-50d	$\gamma : 2.0$	$\beta : 0.9999$	-	-	0.853
C05	ResNeSt-50d	$\gamma : 2.0$	-	th : 0.3	-	0.852
C06	ResNeSt-50d	$\gamma : 2.0$	-	th : 0.5	-	0.844
C07	ResNeSt-50d	$\gamma : 2.0$	-	th : 0.7	-	0.846
C08	ResNeSt-50d	$\gamma : 2.0$	-	-	Multiply the losses of none-classes by 0.1	0.837
C09	ResNeSt-50d	$\gamma : 2.0$	-	-	Multiply the losses of none-classes by 0.01	0.825
C10	ResNeSt-50d	$\gamma : 2.0$	-	-	Snapshot Ensemble of Model-C09	0.823
C11	ResNeSt-50d	$\gamma : 2.0$	-	-	Outlier removal technique	0.843
C12	ResNeSt-50d	$\gamma : 2.0$	-	-	Symmetric Lovasz Loss with square root scaling	0.849
D13	SE-IBN-ResNet50	$\gamma : 2.0$	-	-	-	0.824

(1) **Label Encoding** for reducing label noise and using annotators' information efficiently
(2) **SoTA Architectures (ResNeST/RegNet/HRNet)** for improving baseline performance
(3) **Class-Balanced/Reduced Focal Loss** for tackling imbalanced dataset problems
(4) **Add some tricks** like ignoring outliers / snapshot ensembles / symmetric lovasz loss / etc.

© Hitachi, Ltd. 2020. All rights reserved.

(1) Label Encoding

HITACHI
Inspire the Next

- Use **Label Encoding** for training dataset
  to reduce label noise and use annotators' information efficiently

Original label examples:

```
"DSC_2568.JPG": [
   "['damage:misc', 'damage:rubble']",
   "['damage:none']",
   "['damage:misc', 'damage:rubble']"
],
"DSC_0545_497e3958-100e-4226-801f-0763f83be8da.jpg": [
   "['damage:flood/water']",
   "['damage:none']",
   "['damage:flood/water', 'damage:washout']"
```

Encode by counters

Use these labels as Ground Truth after normalization (clip [0.0, 1.0]) (the length of the list is 37)

DSC_2568.JPG:

label: [2, 0, 0, 0, 2, 0, ···, 1, 0, 0, 0, 0]

misc rubble damage: none

DSC_0545_497e3958···.jpg:

label: [0, 2, 0, 1, 0, 0, ···, 1, 0, 0, 0, 0]

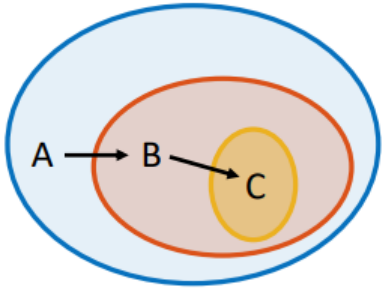
flood/water washout damage: none

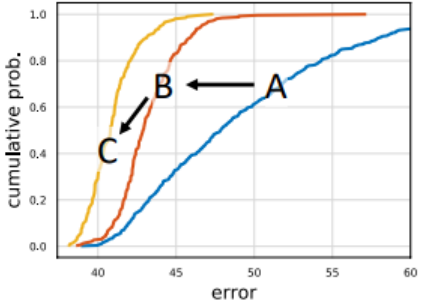
© Hitachi, Ltd. 2020. All rights reserved.

(2) SoTA Architectures (ResNeST/RegNet/HRNet)

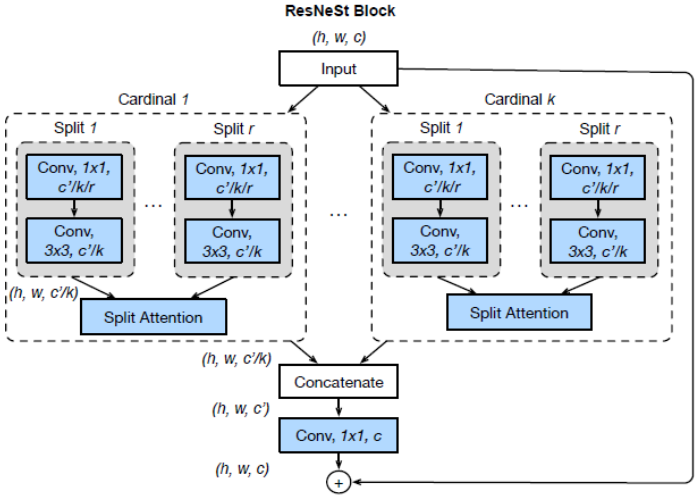
HITACHI
Inspire the Next

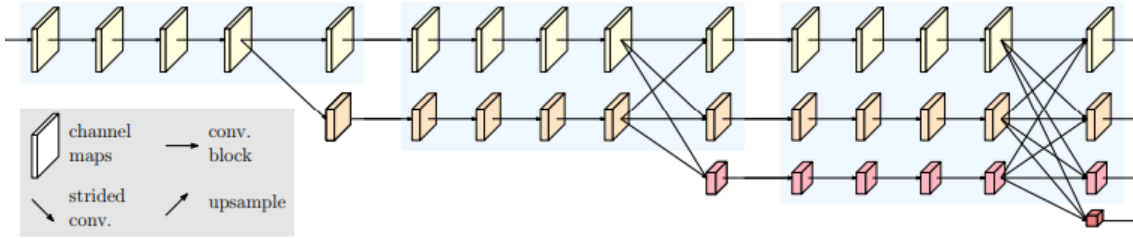
- Use **SoTA architectures (ResNeST/RegNet/HRNet)** as a backbone **to improve baseline performance**.





[RegNet : Arxiv – 2020/03]





[ResNeST : Arxiv – 2020/04]

[HRNetV2 : Arxiv – 2019/04]

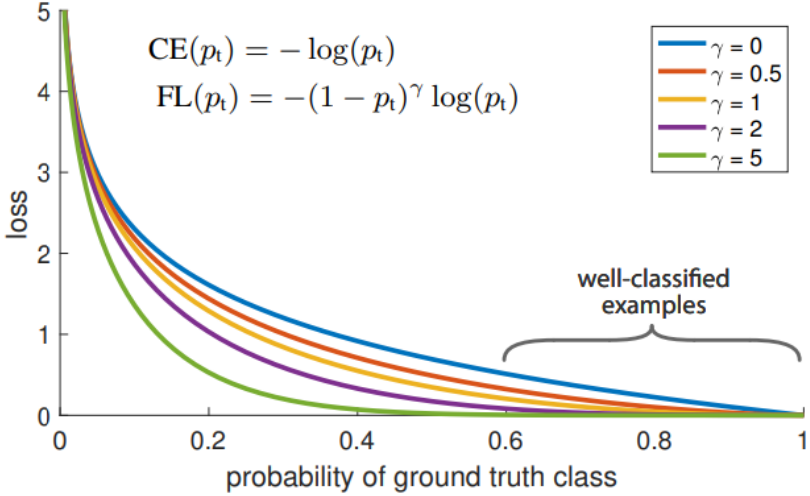
© Hitachi, Ltd. 2020. All rights reserved.

(3) Class-Balanced/ Reduced Focal Loss

HITACHI
Inspire the Next

- Incorporate **Class-Balanced/ Reduced Focal Loss** **to tackle the imbalanced data problem**.

FL (Focal Loss): highly weight the rare classes

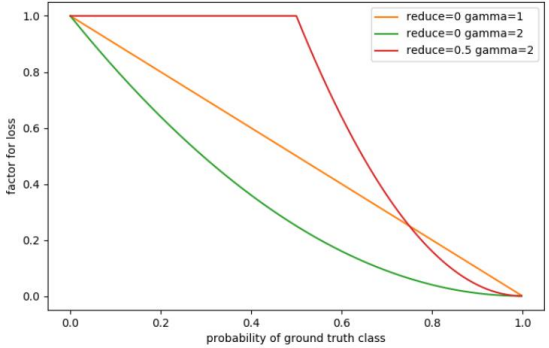


+ Theoretically grounded class-balanced loss using “effective number of samples”.

⇒ **Class-Balanced Focal Loss** [CVPR 2019]

+ softening the loss weight of hard samples and continuing to perform hard example mining

⇒ **Reduced Focal Loss** [Arxiv 2019/03]



© Hitachi, Ltd. 2020. All rights reserved.

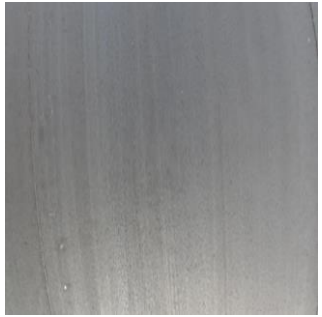
(4) Some Tricks for training

HITACHI
Inspire the Next

- **Ignore max value of losses** in each mini-batch for **treating outliers**

Normal training examples

Outlier training examples (tend to have high loss value)



※ **The length of this list is 28 (equal to mini-batch size)**

losses (before): [0.42, 0.89, 0.23, 0.11, 2.56, 0.12, ···, 0.01, 0.99, 0.21, 0.54, 0.63]

losses (after): [0.42, 0.89, 0.23, 0.11, 0, 0.12, ···, 0.01, 0.99, 0.21, 0.54, 0.63]

© Hitachi, Ltd. 2020. All rights reserved.

(4) Some Tricks for training

HITACHI
Inspire the Next

◻ Multiplying the losses of none-class by small value for classifying named-class efficiently


```
"DSC_2568.JPG": [
    "['damage:misc', 'damage:rubble']",
    "['damage:none']",
    "['damage:misc', 'damage:rubble']"
],
"DSC_0545_497e3958-100e-4226-801f-0763f83be8da.jpg": [
    "['damage:flood/water']",
    "['damage:none']",
    "['damage:flood/water', 'damage:washout']"
```


We don't need to classify none-class correctly for test data, so weight small value to losses of 5 none-class for classifying 32 named-class.

※ The length of this list is 37 (32 named-class + 5 none-class)

losses (before): [0.12, 0.19, 0.13, 0.01, 0.16, 0.42, ···, 0.01, 0.99, 0.21, 0.54, 0.63]

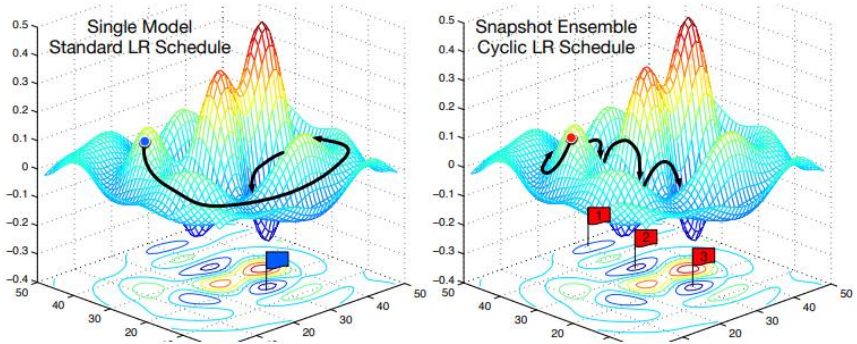
Multiply the loss of none-class by 0.1

losses (after): [0.12, 0.19, 0.13, 0.01, 0.16, 0.42, ···, 0.001, 0.099, 0.021, 0.054, 0.063]

© Hitachi, Ltd. 2020. All rights reserved.

(4) Some Tricks for training

- Use **Snapshot Ensembles** [ICLR 2017] for more ensembling



- Add **Symmetric Lovasz Losses** [CVPR 2018] with square root scaling for creating diversity in ensemble (ref: https://www.kaggle.com/c/tgs-salt-identification-challenge/discussion/69053)

```
loss = loss + torch.sqrt((lovasz_hinge_loss.lovasz_hinge(outputs, targets) + \
                         lovasz_hinge_loss.lovasz_hinge(-outputs, 1-targets)) \
                         / 2.0)
```

© Hitachi, Ltd. 2020. All rights reserved.

Experiment Setting / Metrics for Validation

HITACHI
Inspire the Next

- Architecture: ResNeST-50d / RegNetX-6.4F / HRNet-W18 (model size are almost same as ResNet50)

  Augmentation: Resized to 360*360, (※for HRNet-W18, resized to 448*448)
  RandomHorizontalFlip(p=0.5) / VerticalFlip(p=0.5) / Rotation(degrees=90),
  ColorJitter() (※To shorten inference time, we didn't use Test-Time Augmentation)

  Batch Size: 28 for training, Epoch: 40, Other Hyperparameters : Same as IBN-Net GitHub Repo.

- Metrics: Ranking Loss, Ranking Average Precision, and Precision/Recall/F1-Score by classification report (ref: https://scikit-learn.org/stable/modules/classes.html#module-sklearn.metrics)

In these metrics, ground truth labels' list must have binary-type value, so we further encoded these labels by threshold parameter:

Sample label: [2, 0, 0, 0, 2, 0, ···, 1, 0, 0, 0, 0]

↓

Normalized label: [1.0, 0, 0, 0, 1.0, 0, ···, 0.5, 0, 0, 0, 0]

↓

Clip with th0.6: [1, 0, 0, 0, 1, 0, ···, 0, 0, 0, 0, 0]

Clip with th0.2: [1, 0, 0, 0, 1, 0, ···, 1, 0, 0, 0, 0]

© Hitachi, Ltd. 2020. All rights reserved.

Processing test shot clips / Inference with test images / Ensemble

HITACHI
Inspire the Next

- Create test images from 1825 test shot clips using ffmpeg.

  (※ command: ffmpeg -i sample.mp4 -r 10 -qmin 1 -q 1 -vcodec png ./sample_%04d.png)

  ⇒ With 10fps encoding, we can get 220018 test images.

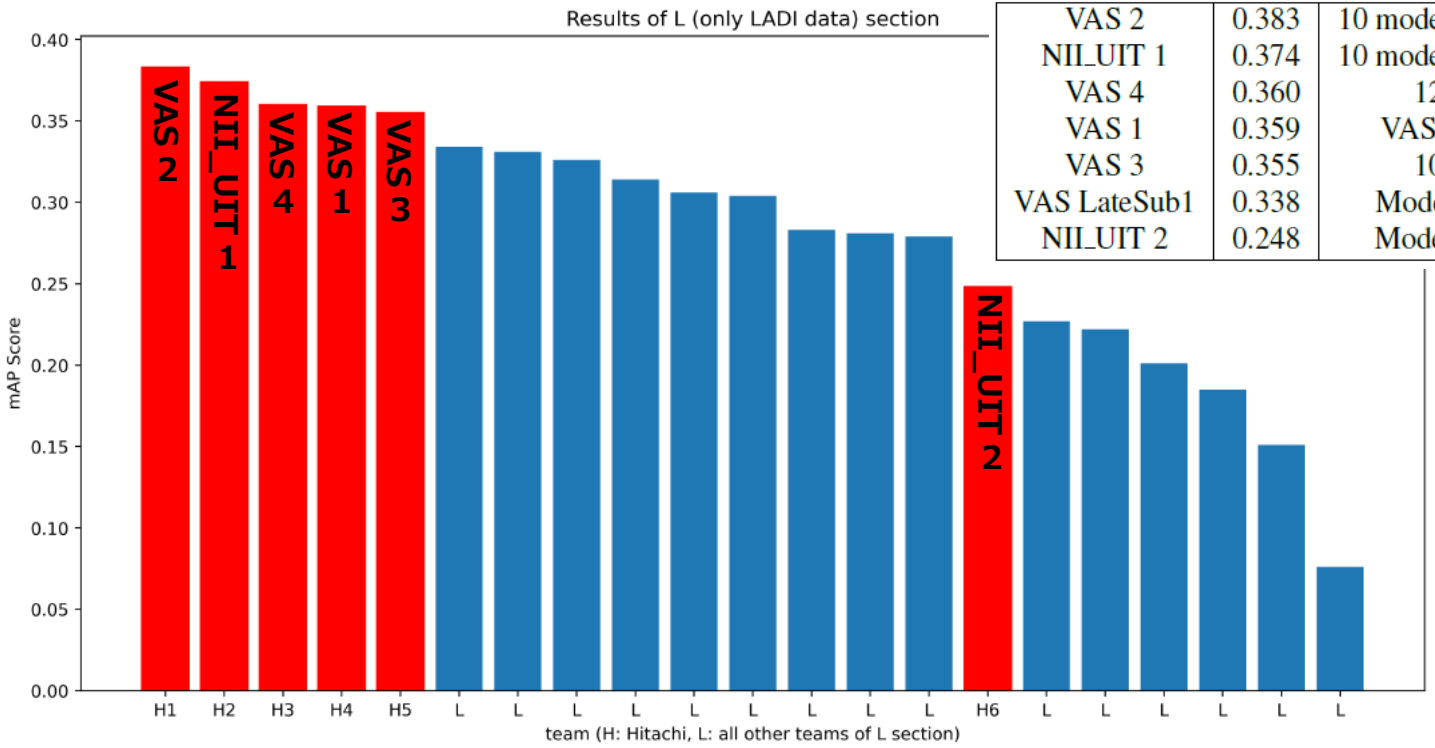
- Use trained models to infer for test images,
  and **extract max value output** in each shot image list as a confidence of each class

	Outputs (Confidences) of washout class	**Outputs (Confidences) of rubble class**
Shot1_001.mp4:	Shot1_001_001.png – 0.225 Shot1_001_002.png – 0.243 Shot1_001_003.png – 0.250 Shot1_001_004.png – **0.576** (max value) …… Shot1_001_134.png – 0.325 Shot1_001_135.png – 0.315	Shot1_001_001.png – 0.125 Shot1_001_002.png – **0.143** (max value) Shot1_001_003.png – 0.050 Shot1_001_004.png – 0.076 …… Shot1_001_134.png – 0.025 Shot1_001_135.png – 0.015

- Ensemble 12 models' confidences with equal weight for each class ⇒ VAS4: mAP-0.360

© Hitachi, Ltd. 2020. All rights reserved.

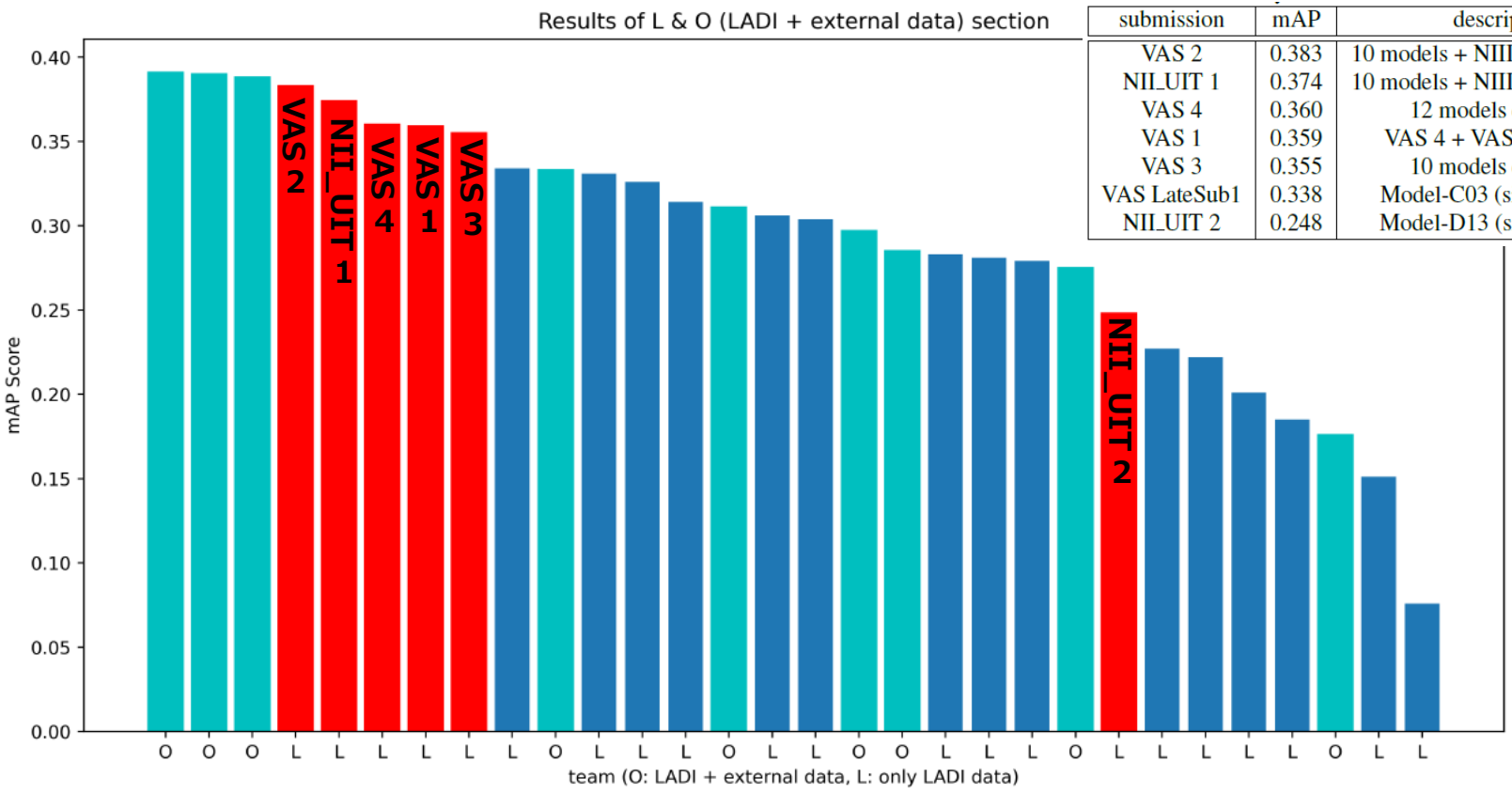
Result of L (only LADI data) section



submission	mAP	description
VAS 2	0.383	10 models + NIIICT model (2:1)
NII_UIT 1	0.374	10 models + NIIICT model (5:1)
VAS 4	0.360	12 models ensemble
VAS 1	0.359	VAS 4 + VAS 3 ensemble
VAS 3	0.355	10 models ensemble
VAS LateSub1	0.338	Model-C03 (single model)
NII_UIT 2	0.248	Model-D13 (single model)

© Hitachi, Ltd. 2020. All rights reserved.

Result of L + O (LADI & external data) section



submission	mAP	description
VAS 2	0.383	10 models + NIIICT model (2:1)
NII_UIT 1	0.374	10 models + NIIICT model (5:1)
VAS 4	0.360	12 models ensemble
VAS 1	0.359	VAS 4 + VAS 3 ensemble
VAS 3	0.355	10 models ensemble
VAS LateSub1	0.338	Model-C03 (single model)
NII_UIT 2	0.248	Model-D13 (single model)

© Hitachi, Ltd. 2020. All rights reserved.

Result

HITACHI
Inspire the Next

[only-internal-data section]

Team	Score
Hitachi w/ NICT Model	**0.383**
Hitachi w/o NICT Model	**0.360**
NICT	0.334
Shield	0.314
VCL	0.283
UTS	0.281
COVIS	0.151
ITI_CERTH	0.076

※Results including
using-external-data section

Team	Training Data	Score
FIU_UM	External	0.391
Hitachi	**Internal**	**0.383**
VCL	External	0.333
Shield	External	0.297

submission	mAP	description
VAS 2	0.383	10 models + NIIICT model (2:1)
NII_UIT 1	0.374	10 models + NIIICT model (5:1)
VAS 4	0.360	12 models ensemble
VAS 1	0.359	VAS 4 + VAS 3 ensemble
VAS 3	0.355	10 models ensemble
VAS LateSub1	0.338	Model-C03 (single model)
NII_UIT 2	0.248	Model-D13 (single model)

© Hitachi, Ltd. 2020. All rights reserved.