CARDS FOR SCIENTISTS FROM THEIR NON-SCIENTIST RELATIONS

Challenges associated with integration of machine vision algorithms into catch accounting programs

an overview of the FMA EM Innovation project

Pacific States Marine Fisheries Commission

Suzanne Romain suzanne.romain@noaa.gov

Craig Rose
raig.rose@noaa.gov

Graeme LeeSon graeme.leeson@noaa.gov

NOAA

Jennifer Ferdinand jennifer.ferdinand@noaa.gov

Paul Packer [paul.packer @noaa.gov](mailto:paul.packer@noaa.gov)

Andy Kingman andy.kingman@noaa.gov

University of Washington
Department of Electrical and Computer Engineering

Jenq-Neng Hwang hwang@uw.edu

Jie Mei jiemei@uw.edu

Aotian Zheng aotianzheng@gmail.com

Summary of EMI development and image types

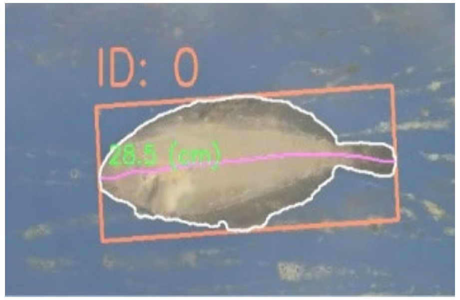
Rail, Camera Chute, Conveyor, Open Deck (Crew on Deck & Fish lengths on tables)

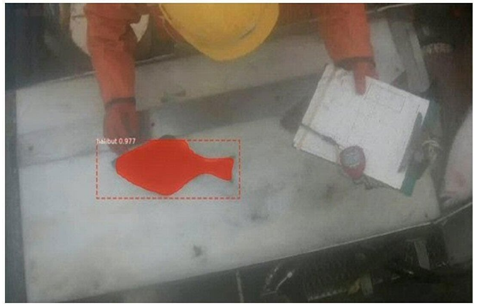
Developed with

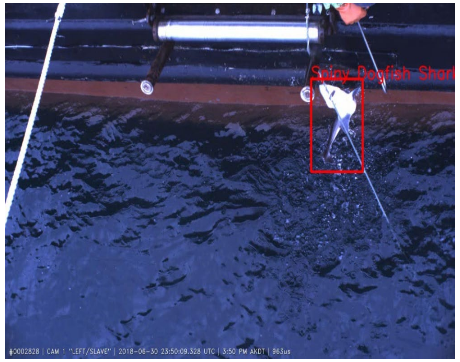
- Controlled lighting/background images
- Outdoor captures of fishing
- Conveyor belts in plants

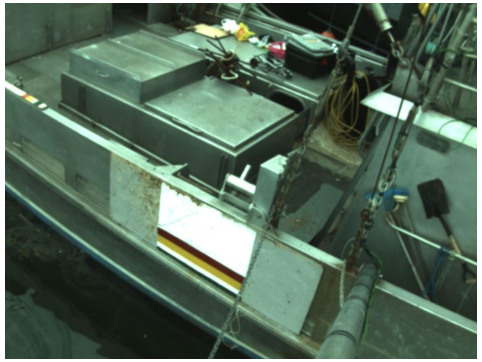
Infer catch accounting information from images

- Hauls
- Species Composition
- Census of species of interest
- Length

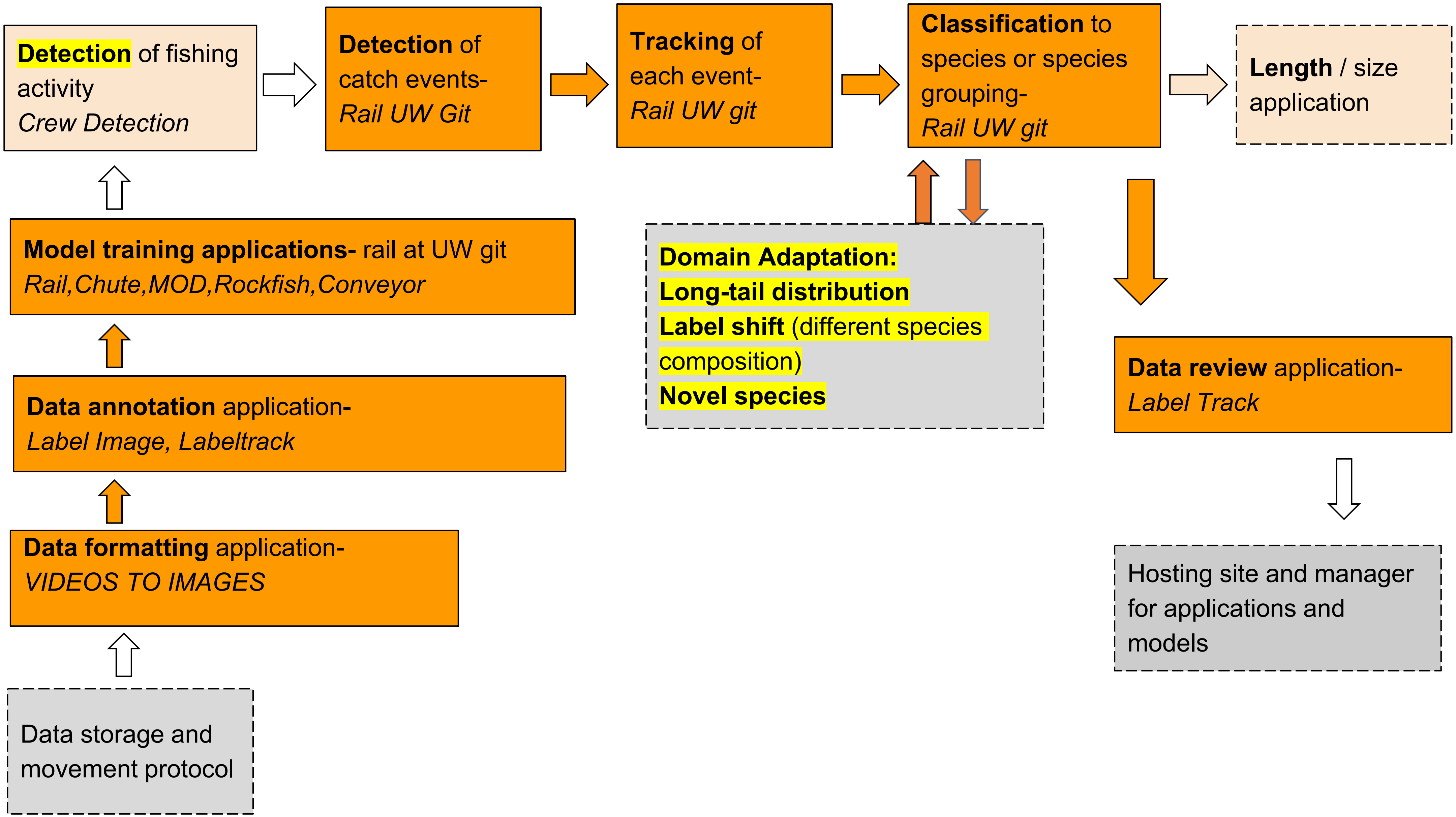








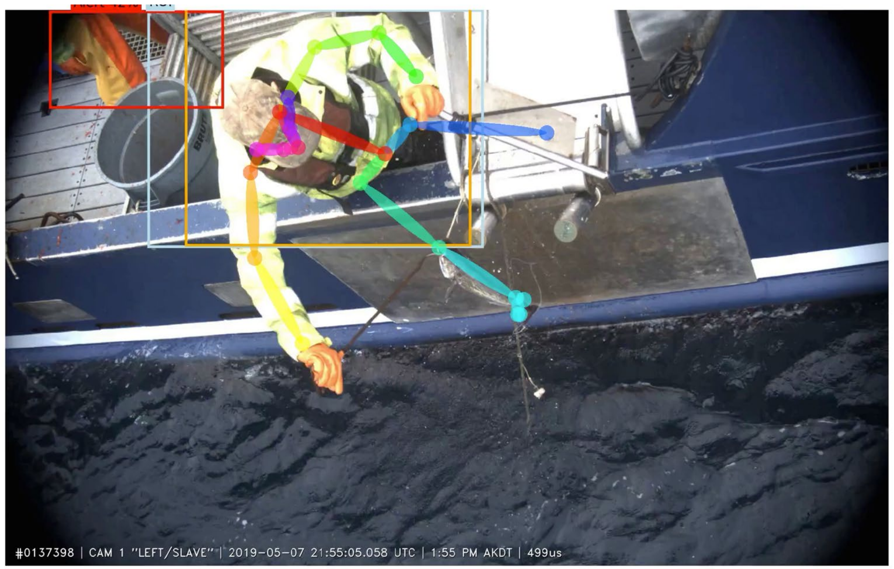
Rail Catch Accounting Pipeline



Machine Vision tool to select data of interest for review

Human Presence Detector & Pose estimation in selected ROI

- 19,903 images from six camera angles were used
- Capable of detecting man on deck at a **98.87%** precision
- Refines identification of hauling periods
- Real-time processing was achievable
- Human pose estimation looks promising
- Requires thoughtful placement of region of interest to capture crew behaviour that can infer specific fishing activities



Yolov3: Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." arXiv preprint arXiv:1804.02767 (2018).

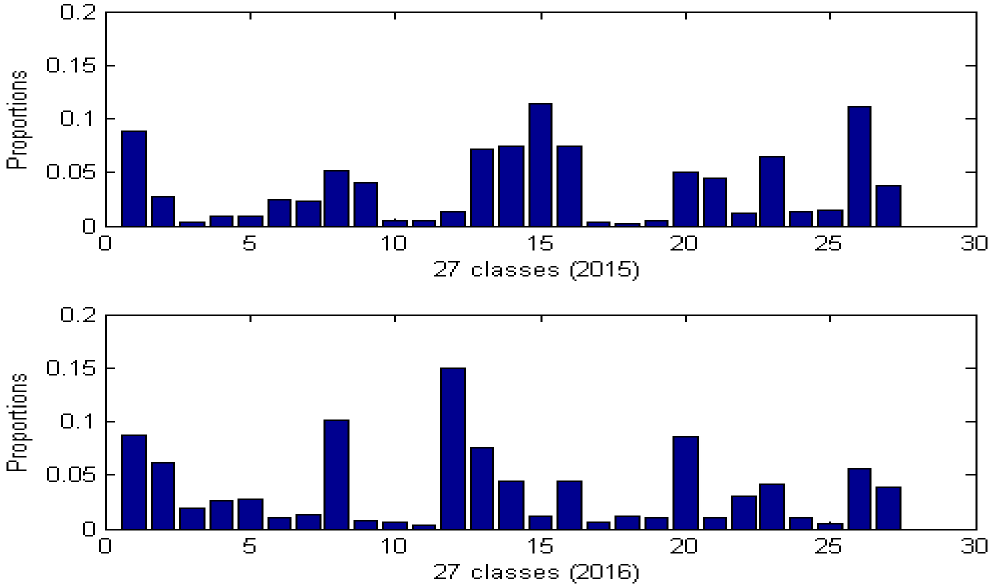
OpenPose [Cao, Z., Hidalgo, G., Simon, T., Wei, S.E. and Sheikh, Y., 2018. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. *arXiv preprint arXiv:1812.08008*.]

Generalized Model…not so much

- Datasets
  - 2015 chute data (8835 images with 27 classes)
  - 2016 chute data (5032 images with 27 classes)

  - Large (domain or label shift) difference between training and testing datasets
    - Slight species variations
    - Different camera color responses
    - Different distributions of species

Changes Species Distributions (Label Shift)



Full Trainings

Training Data	Testing Data	Accuracy (%)
2015	2015	96.1
2016	2016	98.5
2015+2016	2015+2016	96.9

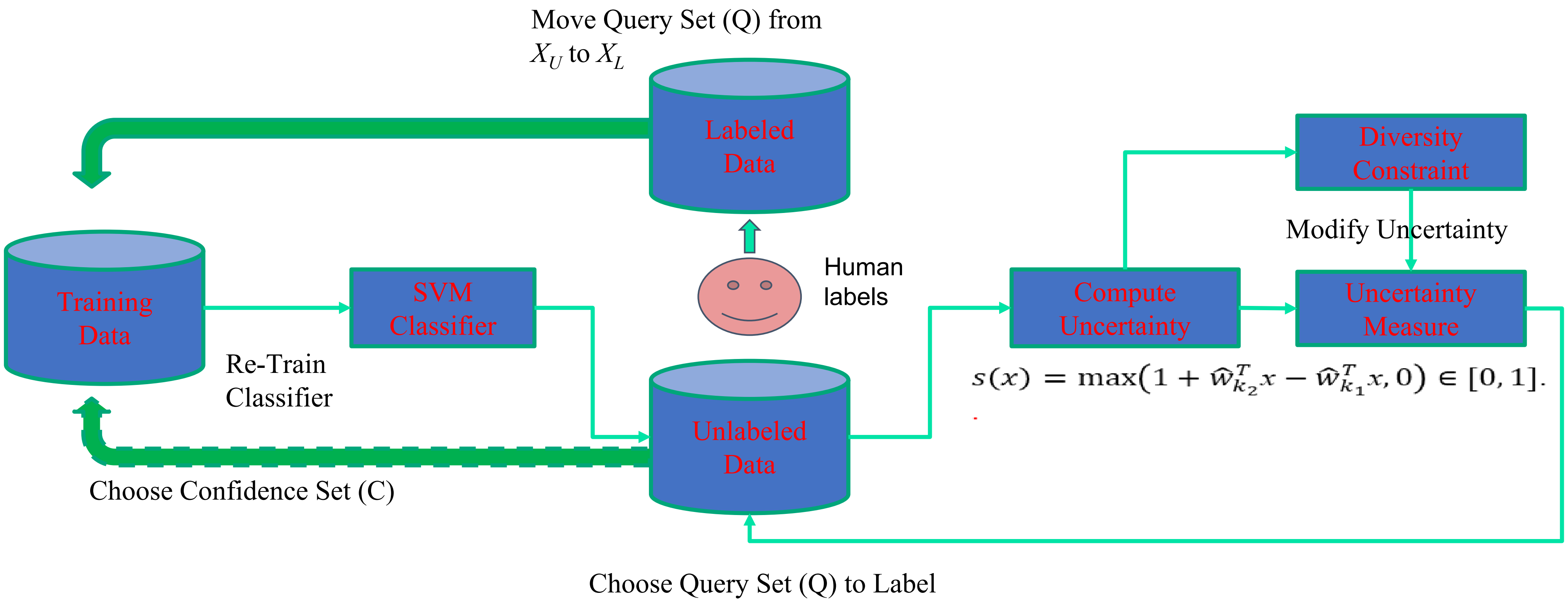
Training with smaller percentages

Training Data	Testing Data	Acc (%)
2015 dataset (5%)	2015 dataset (95%)	83.9
2016 dataset (5%)	2016 dataset (95%)	86.6
2015 dataset (100%)	2016 dataset (100%)	69.5
2015 dataset+2016 dataset (5%)	2016 dataset (95%)	88.1

Query Learning (Semi-Supervised Incremental Learning) to Solve both Domain and Label Shift

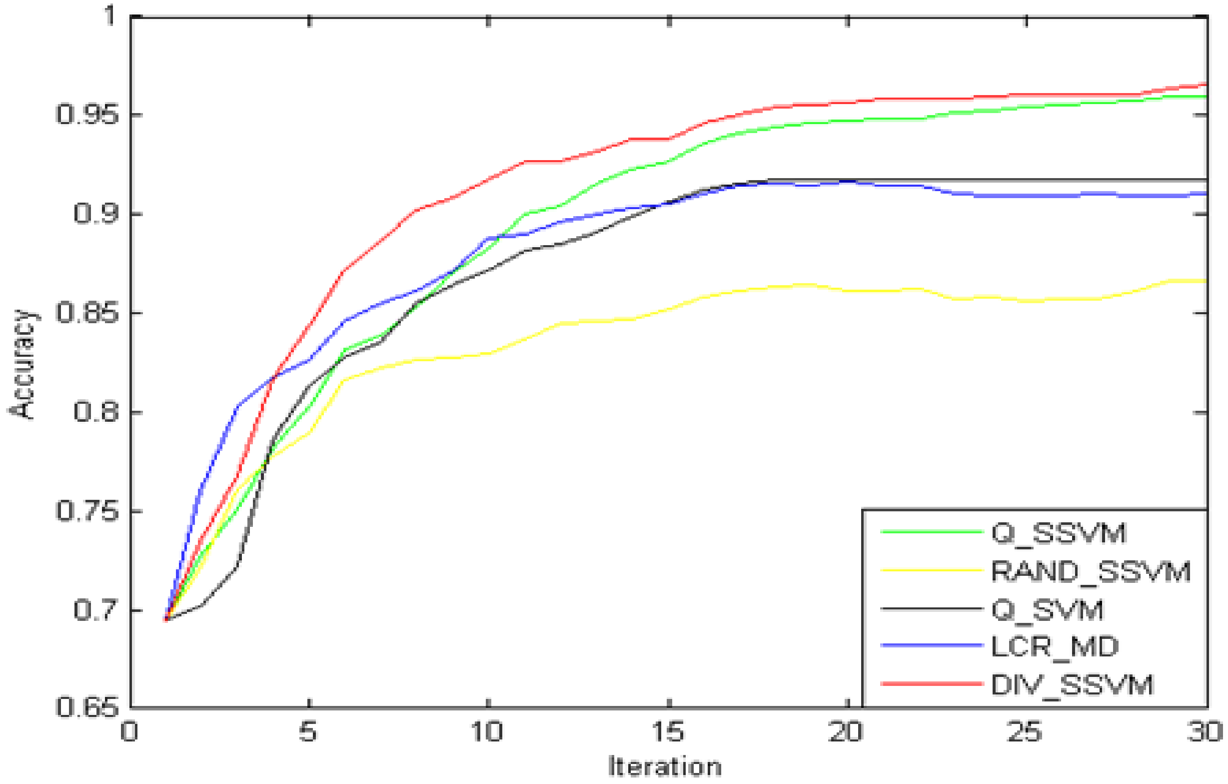
- Requires Human review resources

- Goal: iteratively select informative samples for human labeling to improve the performance of the classifier.



Active Learning for Domain Adaptation

- Training set: 2015 dataset+2016 dataset (5%)
- Accuracy: 96.8% (88.1% to 96.8%)



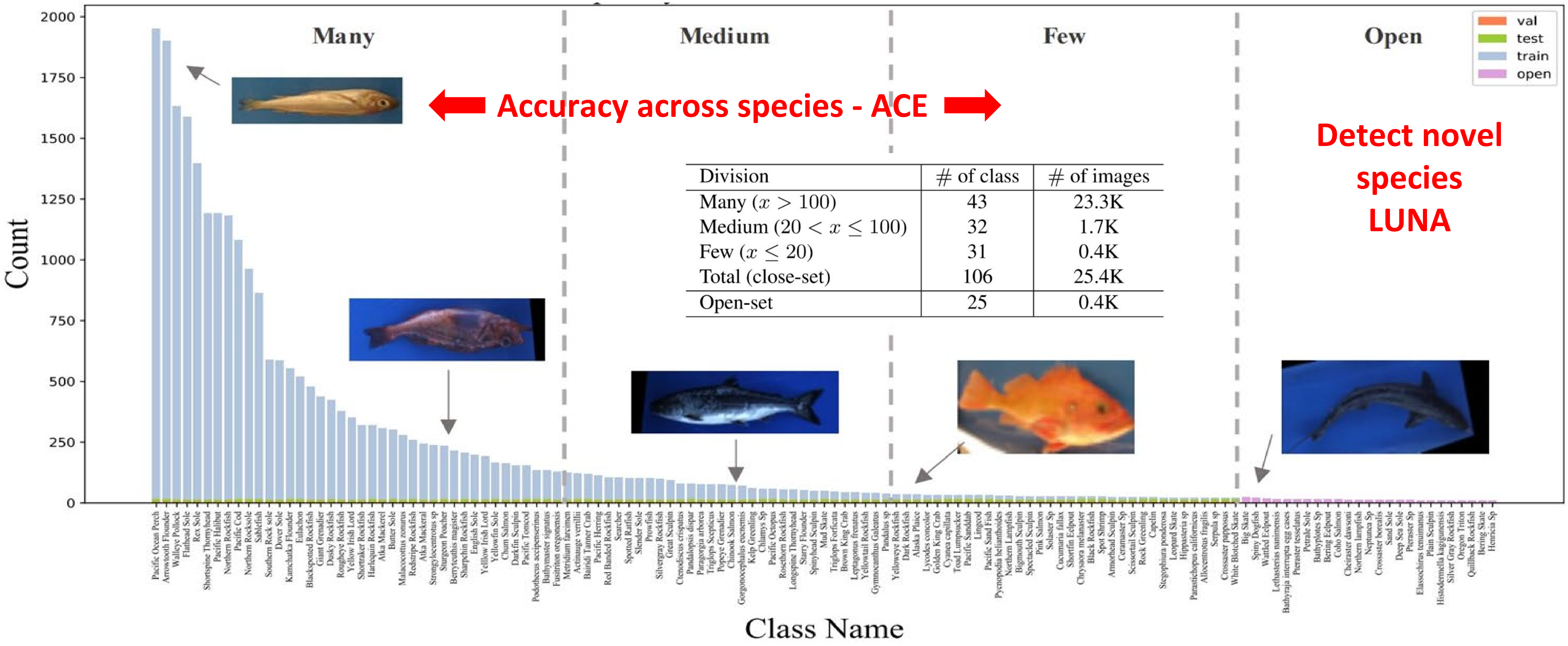
- Q_SSVM: Query learning with semi-supervised learning without diversity constraint.
- RAND_SSVM: Query learning based on random sample selection.
- Q_SVM: Query learning without semi-supervised learning.
- LCR_MD: (Leng et al. 2013).
- DIV_SSVM: Query learning with both semi-supervised learning and diversity constraint.

EMI image collection: Long Tailed Distribution

Avoid Forgetting

Transfer Knowledge

Sensitivity to Novelty



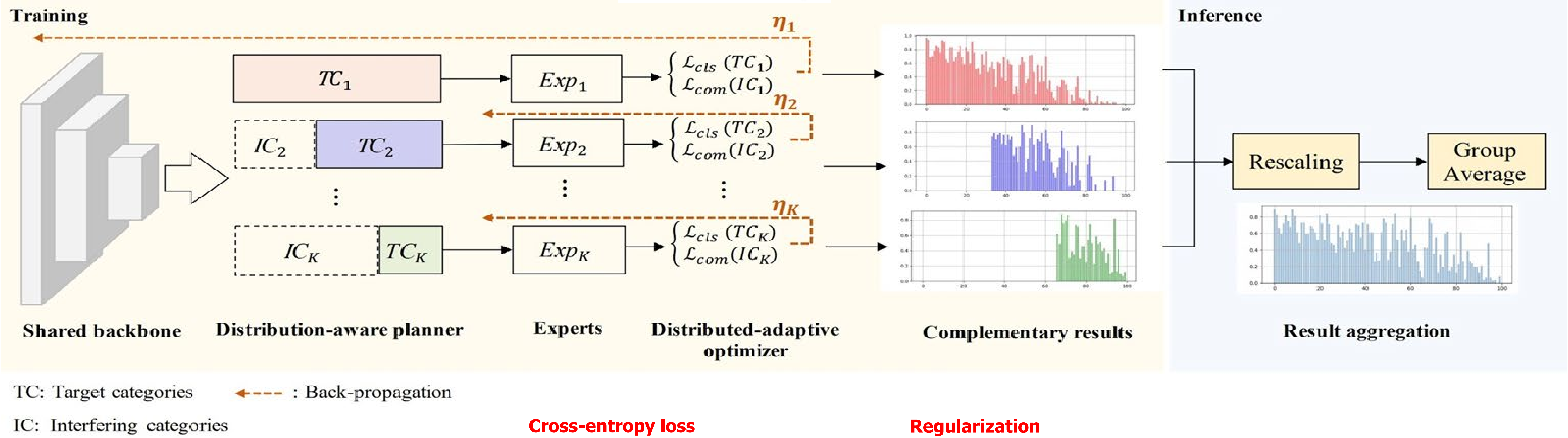
Accuracy across species - ACE

Detect novel species LUNA

Division	# of class	# of images
Many ($x > 100$)	43	23.3K
Medium ($20 < x \leq 100$)	32	1.7K
Few ($x \leq 20$)	31	0.4K
Total (close-set)	106	25.4K
Open-set	25	0.4K

Domain Adaptation for long tailed distributions across all classes in one shot: ACE: Ally Complementary Experts

$$\eta_i = \eta_0 \cdot \frac{\sum_{c \in \mathcal{C}_i} n_c}{\sum^{\mathcal{C}} n_j}$$



Cross-entropy loss

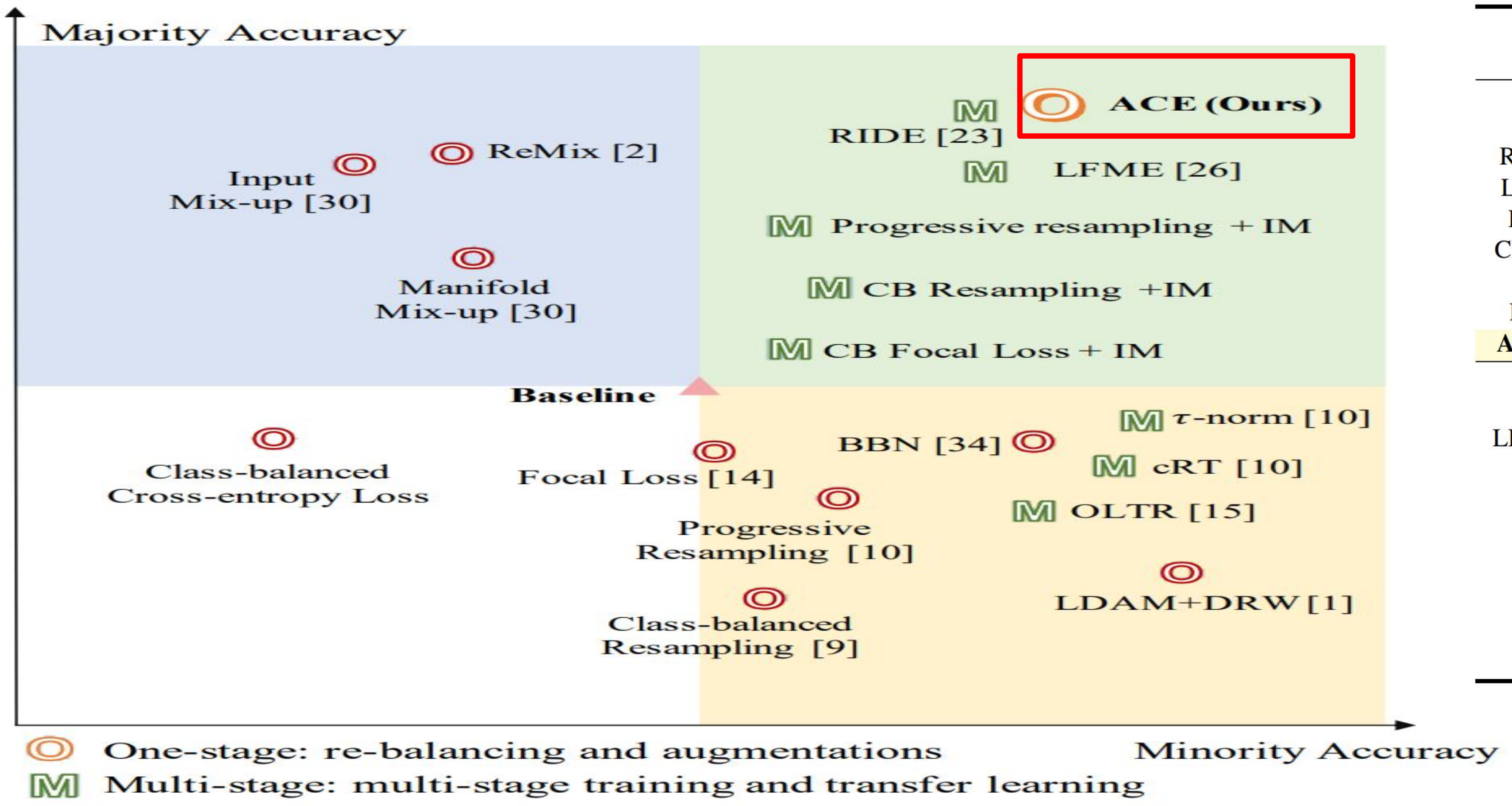
Regularization

$$L^i = L^i_{cls} + L^i_{com} = -\sum^{\{TC_i\}} y\log(\sigma(z_i)) + \sum_{c_j}^{\{IC_i\}} \left\| z_i^{c_j} \right\|^2$$

Jiarui Cai, et al., “ACE: Ally Complementary Experts for Solving Long-Tailed Recognition in One-Shot,” ICCV 2021

Domain Adaptation across all classes in one shot: ACE: Ally Complementary Experts

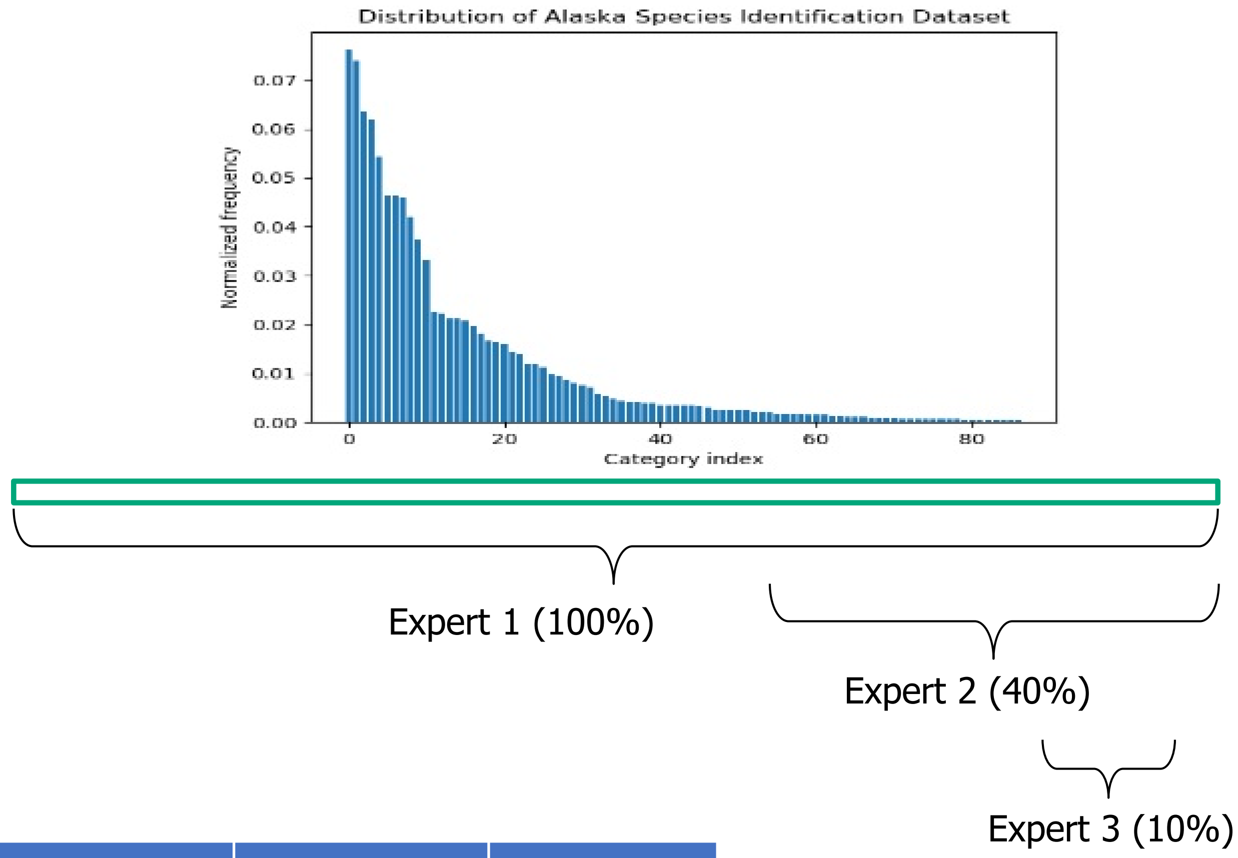
Performance



Method	ImageNet-LT			iNaturalist
	Res10	Res50	ResX50	Res50
Baseline	20.9	41.6	44.4	66.1
FSLwF [5]	28.4	-	-	-
Range Loss [31]	30.7	-	-	-
Lifted Loss [18]	30.8	-	-	-
Focal loss [14]	30.5	-	-	60.3
CB Focal loss [3]	-	-	-	61.1
BBN [34]	-	48.3	49.3	68.0
Logit Adj.[16]	-	51.1	-	66.4
ACE (3 experts)	**44.0**	**54.7**	**56.6**	**72.9**
OLTR [15]	34.1	-	46.3	63.9
NCM [10]	35.5	44.3	47.3	-
LDAM+DRW [1]	36.0	-	-	68.0
cRT [10]	41.8	47.3	49.5	65.2
τ-norm [10]	40.6	46.7	49.4	65.6
LWS [10]	41.4	47.7	49.9	65.9
CAM [32]	43.1	-	-	70.9
LFME [26]	38.8	-	-	-
RIDE [23]†	-	54.4	55.9	71.4
RIDE [23]‡	-	54.9	56.4	72.2

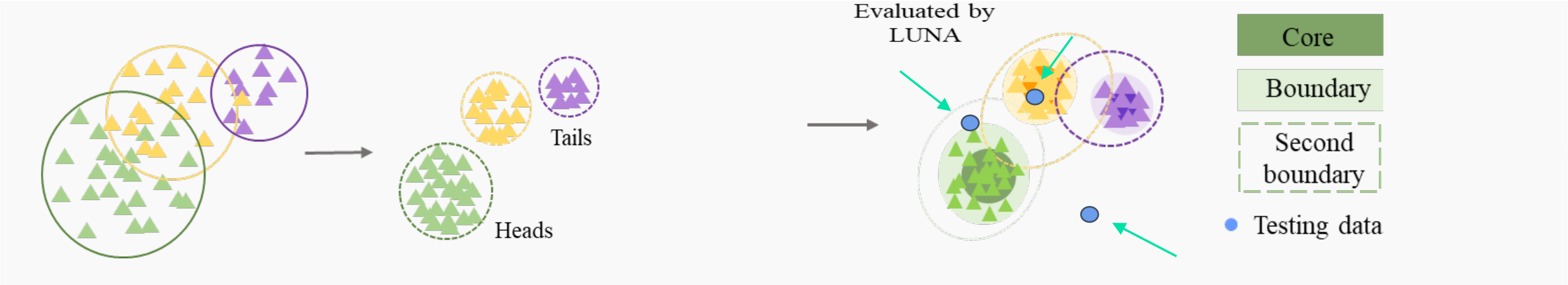
Domain Adaptation across all classes in one shot: ACE: Ally Complementary Experts : Performance

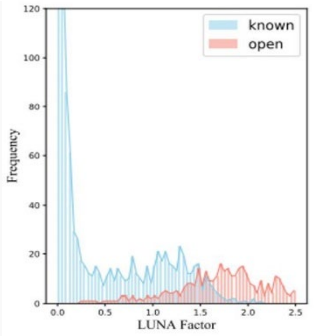
- Alaska species ID dataset: 26.4k images for 87 classes
- Many-shot (>100 samples): 38 classes
- Medium-shot (>20 and <=100 samples): 33 classes
- Few-shot (<= 20 samples): 16 classes
- Imbalance factor = Nmax/Nmin= 193.5
- Backbone: ResNet 50
- Number of experts: 3
- Experts 1, 2, 3 are trained with 100%, 40% and 10% of the dataset, respectively.

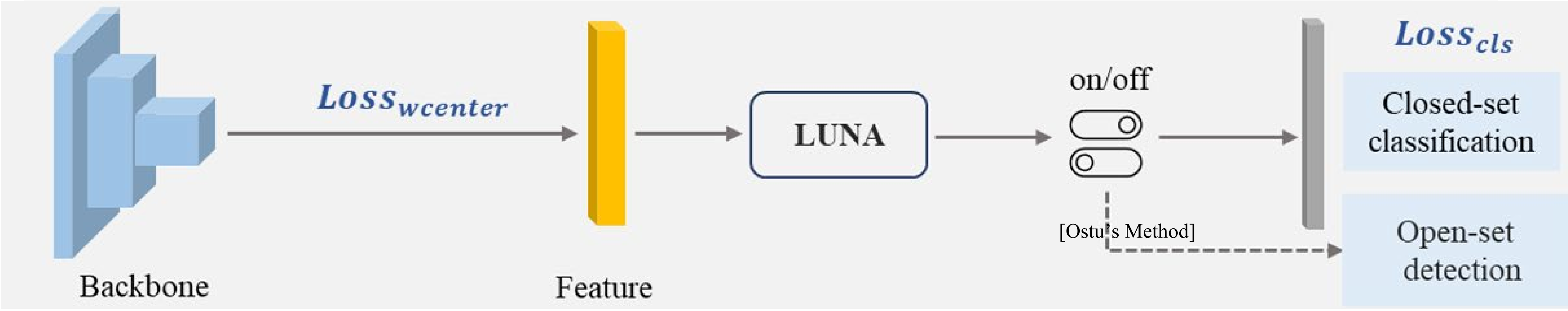


	Overall Accuracy	Many	Medium	Few
ACE	95%	98%	97%	80%

Localizing Unfamiliarity Near Acquaintance (LUNA)



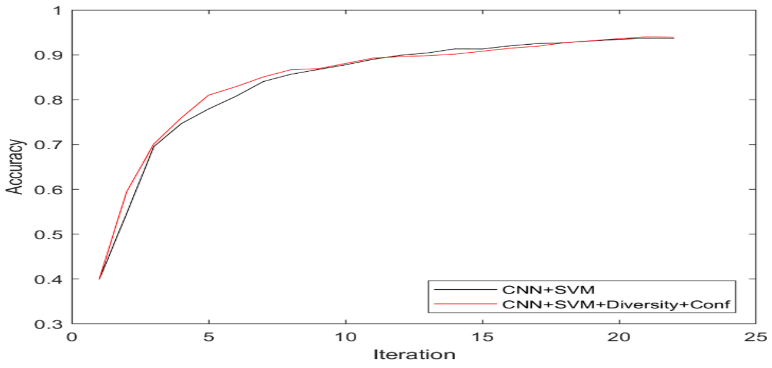




- Non-Query Learning
- 43-class (42+1 others)
  - Training, 6042 images
  - Testing, 698 images
  - 90% samples used in the training.
  - Accuracy = 94.5%.

- New Class Discovery
- From 27 to 42 classes
  - 5% samples used in the training.
  - Accuracy = 93.9%.

Jiarui Cai, et al., 2022



Maintaining Classification Accuracy: Domain Adaptation

Development image types

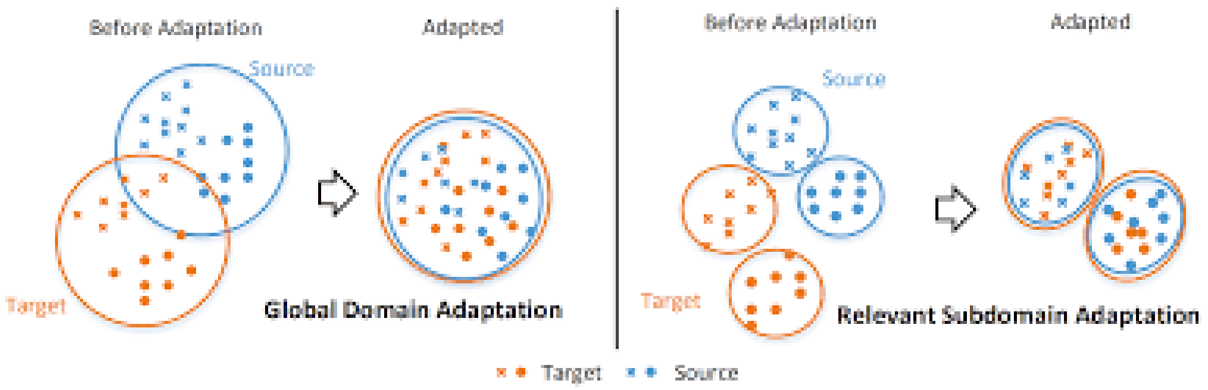
New data image types

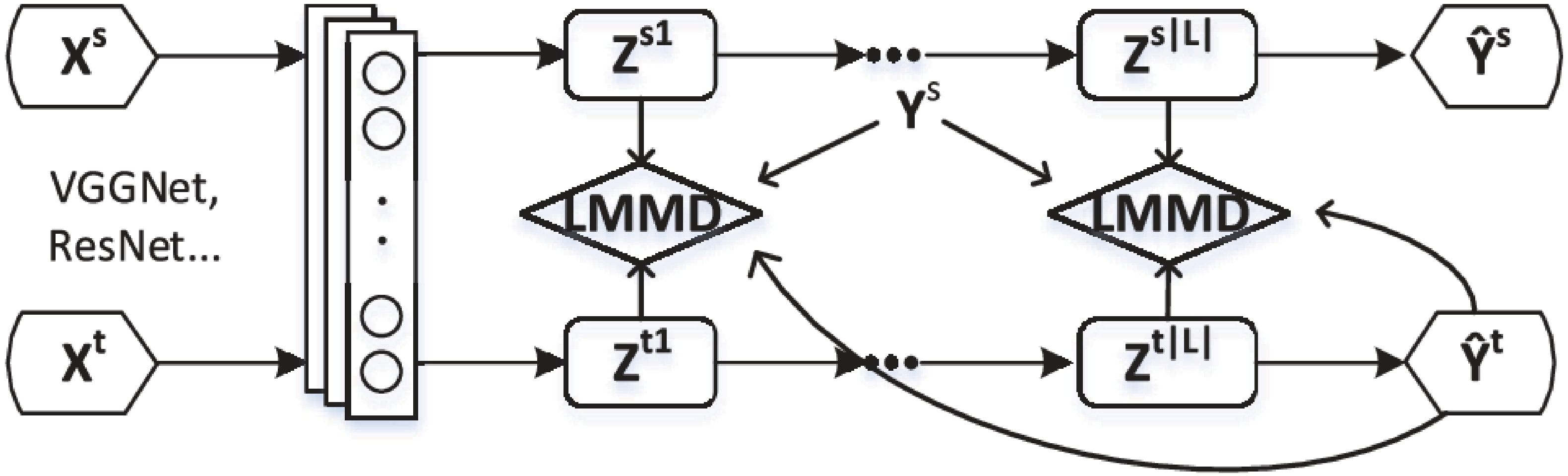
- Needed to address Changes to:
  - Camera angles
  - Background
  - Lighting conditions
  - Species composition
- Can be Unsupervised or Supervised, depending on human review resources
- Natural distribution of our image data challenges classification
- No "none of the above" option for machine vision classifiers, novel species need a specific strategy

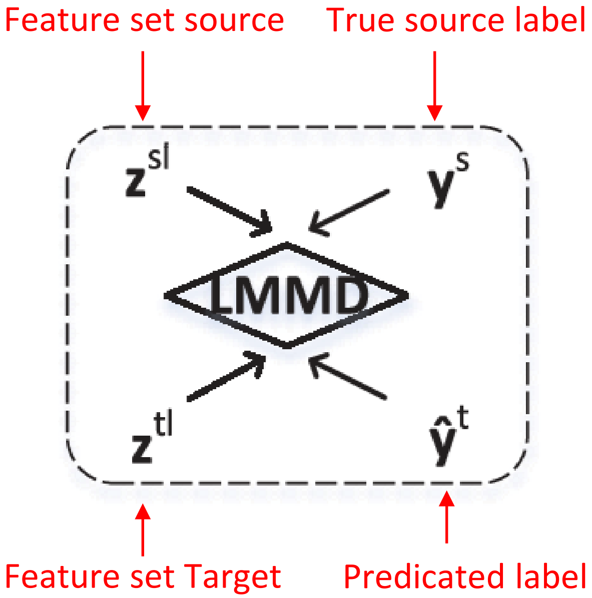
Deep Subdomain Adaptation Network for Image Classification

Local Maximum Mean Discrepancy (LMMD)

- Uses labeled source data and unlabeled target data
- Extends the feature representation by aligning relative sub domains (IE; groups a halibut observation in source domain to another halibut observation in the target domain)







Yongchun Zhu, et al., “Deep Subdomain Adaptation Network for Image Classification” 2021 : IEEE

Data Utilized

labeled image track data from base algorithm training

unlabeled image data from new vessels

Training

Step-1: Pretrain Domain Adaptation universal (all views) and angle/vessel specific algorithms on labeled research ship data

Step-2: Fine tuning with LMMD on labeled research ship data and unlabeled target ship

A: Separately on different target ships

B: Jointly on combined target ships

Evaluation

Using specific adapted models

Using one generic adapted model

Performance

Flat Classifier	# tracks	#frames	# species	No UDA	Generic UDA	Separate UDA
ForwardRollSV_2	87	7,136	10	24.14	57.47	**63.22**
ForwardRollSV_2	108		9	61.11	78.70	**83.33**
AftRollSV_1	90	6,205	7	26.67	95.56	**96.67**
AftRollSV_1	95		8	24.21	96.84	**97.89**
AftRollSV_2	13	3,926	4	23.08	**92.31**	84.62
AftRollSV_2	55		6	1.81	72.73	**85.45**
AftRollSV_2	22		4	31.82	86.36	**95.45**
ForwardRollSV_2	73	1,665	9	53.42	87.67	**90.41**
ForwardRollSV_1	40	1,148	5	32.50	**67.50**	65.00
Top Down_V1	72	2,115	6	6.94	45.83	**54.17**

Implementation of Machine Vision Tools into a catch accounting program

While some applications can be generalized, data needs human configuration of regions of interest and vessel specific adaptions to extract useful information from the outputs

Multiple domain adaptation steps may be needed

Some image sets/types may not be viable with any domain adaptation strategy

Integrating machine vision tools will require

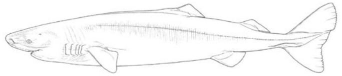
- Image type specific database and processing pipeline
- Configuration parameters for each vessel
- Domain adaptation strategy-could vary based on vessel or target species
- Review/ Audit protocols and staff
- Advance management of computing resources/balancing CPU GPU

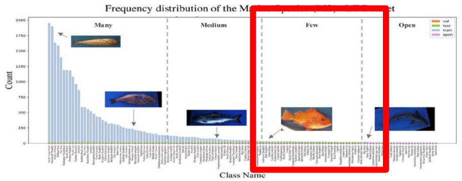
Top Down

Flat Classifier	# tracks	# species	No UDA	Generic UDA	Separate UDA	Gt tracks
Top Down_V1	72	6	6.94	45.83	**54.17**	yes

Rail : Rare Species Data Mining Pipeline

Pacific Sleeper Shark





Data storage and movement protocol

Hosting site manager for applications and models

!?!

- Manage video and image input/output
- Model training set, Vessel configurations
- Optimize CPU/GPU performance (multithreading)
- User interface for database managers and auditors

Extract video files

Vessel Configuration

- **Vessel Name**
- **Trip Dates**
- **FOV image**
- **ROI Crew & Catch**

Human audit

GPS
Hydraulic
sensors

Hauls | **Species Composition** | **Length**

Convert to jpgs

Fishing gear
Bait
Bird (on line or in air)
Fish

Count of each individual catch

Group / Species
Flatfish Roundfish
Rockfish Shark
Skate Invertebrate

Could be multiple applications

Select % for Data review

Detection of fishing activity *Crew Detection*

Detection of catch events

Tracking of catch event

Classification species/group

Domain Adaptation

Select % for refinement

Length / size application

Cai, J., Y. Wang, J.N. Hwang. 2021. Ace: Ally complementary experts for solving long-tailed recognition in one-shot. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pages 112–121

Cai, J., Y. Wang, C. Rose, K. Magrane, J.N. Hwang. 2022. LUNA: Localizing Unfamiliarity Near Acquaintance for Open-Set Long-Tailed Recognition Computer Science AAAI Conference on Artificial Intelligence

Cai, J 2022. Toward Visual Recognition in the Wild. University of Washington http://hdl.handle.net/1773/48785

Fitzgerald, S., F. Wallace, K. Magrane. 2019. Improving seabird species identification in electronic monitoring applications using machine learning systems. In: ACAP - Ninth Meeting of the Seabird Bycatch Working Group. ACAP SBWG9 Inf 21, Florianópolis, Brazil.

Huang, T. J. Hwang, and C. Rose, "Chute based automated fish length measurement and water drop detection," in 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2016, pp. 1906–1910. 1

Huang, T., J. Hwang, S. Romain and F. Wallace. 2019. Fish tracking and segmentation from stereo videos on the wild sea surface for electronic monitoring of rail fishing. *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 10, pp. 3146-3158, Oct. 2019, doi: 10.1109/TCSVT.2018.2872575.

Huang, T., J. Hwang and C. S. Rose. 2016. Chute based automated fish length measurement and water drop detection. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, 2016. pp. 1906-1910, doi: 10.1109/ICASSP.2016.7472008.

Huang, T., J. Hwang, S. Romain and F. Wallace. 2019. Recognizing fish species captured live on wild sea surface in videos by deep metric learning with a temporal constraint., *IEEE International Conference on Image Processing (ICIP)*, Taipei, Taiwan, 2019. pp. 3407-3411, doi: 10.1109/ICIP.2019.8803592.

Huang, T. 2019. Automatic Video Analysis for Electronic Monitoring of Fishery Activities. (Accession No. 2019-10-15T22:57:37Z) Doctoral dissertation, University of Washington. http://hdl.handle.net/1773/44796

Mei, J., J. Hwang, Suzanne Romain, Craig Rose, Braden Moore, and Kelsey Magrane, "Video based hierarchical species classification for longline fishing monitoring," 2021. 1

Mei, J., J.Yu, J. Hwang. S. Romain, C. Rose, K. Magrane, G. LeeSon. 2022 Unsupervised Severely Deformed Mesh Reconstruction (DMR) From A Single-View Image for Longline Fishing.Computer Science 2022 IEEE International Conference on Multimedia and Expo Workshops

Mei, J. Jenq-Neng Hwang, S. Romain, C. Rose, B. Moore, and K.Magrane, 2021. "Absolute 3d pose estimation and length measurement of severely deformed fish from monocular videos in longline fishing," Environmental Science ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

Mei, J. S. Romain, Craig S. Rose, Kelsey Magrane, Jenq-Neng Hwang. 2022 HCIL: Hierarchical Class Incremental Learning for Longline Fishing Visual Monitoring. Environmental Science, Computer Science. International Conference on Information Photonics

Wang, G., 2019. Vision based analysis in fisheries applications Doctoral dissertation, University of Washington. http://hdl.handle.net/1773/43977

Wang, G., J. Hwang, C. Rose and F. Wallace. 2019. Uncertainty-based active learning via sparse modeling for image classification. *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 316-329, Jan. 2019, doi: 10.1109/TIP.2018.2867913.

Wang, G., J. Hwang, Y. Xu, F. Wallace and C. S. Rose. 2018. Coarse-to-fine segmentation refinement and missing shape recovery for halibut fish. *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Anaheim, CA, USA, 2018, pp. 370-374, doi: 10.1109/GlobalSIP.2018.8646442.