Weakly Labeling the Antarctic: The Penguin Colony Case

Hieu Le[1,3], Bento Gonçalves[2,3], Dimitris Samaras[1], and Heather Lynch[2,3]

[1]Department of Computer Science, Stony Brook University
[2]Department of Ecology and Evolution, Stony Brook University
[3]Institute for Advanced Computational Science, Stony Brook University

Abstract

Antarctic penguins are important ecological indicators – especially in the face of climate change. In this work, we present a deep learning based model for semantic segmentation of Adélie penguin colonies in high-resolution satellite imagery. To train our segmentation models, we take advantage of the Penguin Colony Dataset: a unique dataset with 2044 georeferenced cropped images from 193 Adélie penguin colonies in Antarctica. In the face of a scarcity of pixel-level annotation masks, we propose a weakly-supervised framework to effectively learn a segmentation model from weak labels. We use a classification network to filter out data unsuitable for the segmentation network. This segmentation network is trained with a specific loss function, based on the average activation, to effectively learn from the data with the weakly-annotated labels. Our experiments show that adding weakly-annotated training examples significantly improves segmentation performance, increasing the mean Intersection-over-Union from 42.3 to 60.0% on the Penguin Colony Dataset.

1. Introduction

The vast and growing catalogs of high-resolution earth observation imagery present us with unprecedented opportunities for understanding ecological and geological processes, but time and domain expertise are in high demand and building a training dataset sufficient for deep learning is often not feasible. Fortunately, many earth observation applications benefit from dynamics that are slow relative to the repeat frequency of the available imagery. As a result, each image is similar to previous imagery, and prior information in the form of lower resolution or auxiliary information can be used to greatly improve classification accuracy. The use of prior knowledge naturally extends to the classification of imagery time series, which in aggregate can be used to understand the dynamics of landscape change.

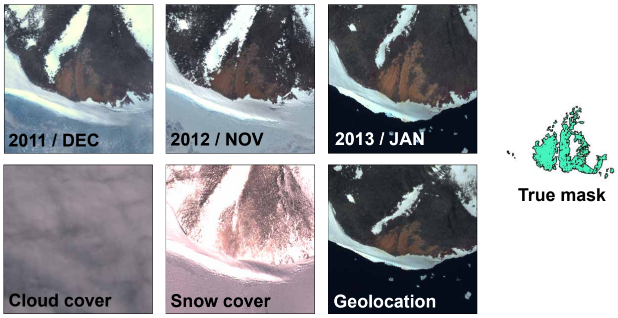


Figure 1. **Penguin colony guano mask extraction on high-resolution satellite imagery.** In the upper row, we show high-quality crops for three consecutive years of the Cape Crozier Adélie penguin colony during the breeding season. Brown to red shapes at the center of crops show the guano stains associated with the breeding colony that, when converted to area, can be used to approximate breeding population size. Boxes in the lower row show unsuitable images cropped at the same location, suffering from occlusion by clouds, heavy snow, and gross orthorectification artifacts. The goal of this work is to use weakly-annotated data, in the form of percent guano coverage, to generate better segmentation masks. Imagery copyright DigitalGlobe, Inc. 2019.

High-resolution satellite imagery is becoming an efficient means to survey inaccessible, or perilous, regions remotely (e.g., [3, 19]). However, real-world applications for semantic segmentation often lack pixel-wise annotations because generating them is so time consuming. In this work we present a framework for weakly-supervised segmentation on georeferenced datasets. Our approach circumvents data acquisition limitations by using pixel-wise mismatched masks, unsuitable by themselves as segmentation groundtruth, to improve segmentation results. Such derived quantities are more robust to geolocation errors than pixel-wise segmentation masks and are easily acquired by a spatial query for available satellite imagery. We demonstrate our approach by segmenting penguin colonies visible in high-resolution satellite imagery, but our method is more broadly applicable to high-resolution segmentation problems common in satellite image analysis.

Antarctic penguins, being sensitive to climate-change driven shifts in the environment [2, 8] and amenable to satellite-based surveys [12], are ideal ecological indicators for the Southern Ocean ecosystem. Brush-tailed penguins, a group of three species in the genus *Pygoscelis*, nest together in large colonies on snow-free rock outcrops along the Antarctic coastline. During the austral summer penguin breeding season, colonies create large-scale guano stains that are visible from space [19] and with an area proportional to breeding population size [12]. In addition to snapshots of population size, we can take advantage of the site-fidelity of penguins to extract time-series of population change by repeatedly surveying colonies via satellite. Time-series of guano stain shape and areal extent are invaluable to furthering our understanding of penguin population dynamics – especially relevant in the face of climate change [6].

Previous satellite-based brush-tailed penguin surveys relied on manual annotation from domain experts [12, 19] or, when automated, suffered from poor transferability between images [30]. Despite their limitations, satellite-based surveys have been used successfully in a variety of contexts, from finding new penguin super-colonies [4] to facilitating the first global population database for Adélie penguins (*Pygoscelis adeliae*) [18]. Manually annotating a single penguin colony, however, takes at least 30 minutes and often significantly longer. Such a laborious process, coupled with the large number of images needed to attain sufficiently temporal depth for time-series analyses at the pan-Antarctic scale, creates an urgent need for robust, automated approaches.

To amass data to train an automated guano extraction tool, we use a small number of hand-drawn geolocated guano polygons [9, 18] as guides to query our high-resolution imagery for images containing penguin colonies. The two issues with this type of weakly-labeled data are: 1) This data-extraction routine is error-prone, potentially generating training images where the corresponding guano stain is not visible, and 2) Re-purposed segmentation masks are imprecise since penguin colonies keep evolving overtime and the images of the same colony are not correctly registered at the pixel level (Fig. 1).

In this paper, we propose a semi-supervised learning framework and a specific loss function to train a segmentation network from a small set of human-annotated images and the weakly-labeled data. We first train a classifier network, C-Net, to verify if an image contains visible guano stains. After being trained, the C-Net can be used to filter out unsuitable training images from the weakly-labeled training set, removing images that were covered in snow, shadows, clouds, or simply were not captured during the penguin breeding season. We demonstrate that, even with a small set of human-annotated images, C-Net successfully weeds out a large proportion of potentially misleading weakly-labeled training images. We employ a C-Net-filtered weakly-labeled training set, combined with our small set of human-annotated images, to train a segmentation network, S-Net, for penguin guano segmentation, using a specific loss function: For hand-labeled images, S-Net is trained to predict pixel-wise guano masks, whereas for weakly-labeled images, S-Net is trained to match percent cover. Our framework with C-Net and S-Net addresses two challenges: 1) how to determine if an image whose coordinates overlap with a penguin colony contains visible guano; and 2) how to use misaligned masks to train segmentation models.

The main contributions of this paper are three-fold: **1)** We propose a data acquisition scheme for georeferenced images, showcased with the Penguin Colony Dataset. From a few hand-drawn segmentation masks, our scheme generates thousands of weakly-labelled training images. **2)** We propose a weakly-supervised framework using a specific loss function to learn segmentation masks from weak labels, in the form of percent cover, and a classifier network to filter out bad training examples. This approach is easily extensible to other georeferenced datasets for segmentation. **3)** We present test results showing that predictions from our framework are superior to pure semantic segmentation approaches and two other baselines across a range of settings.

2. Related Works

The expert annotation labor required to produce segmentation masks hinders the feasibility of fully-supervised deep learning methods. Hence, many deep learning based segmentation work focus on learning from more easily obtainable, weakly-supervised, or synthetic data [5, 15, 16, 17, 28, 32, 26, 14]. A typical example of weak supervision is applying bounding boxes to learn segmentation masks [7, 22]. Some methods can improve segmentation results by learning from as little as a few strokes [27, 29] or points [21, 25]. Malkin *et al.*[20] propose the adoption of statistical descriptors, in the form of the means and variances of low-level annotation masks, to train segmentation networks for high-resolution imagery.

3. Penguin Colony Dataset

We present the Penguin Colony Dataset, a dataset for penguin guano segmentation on high-resolution satellite imagery. Our dataset includes a set of 31 hand-labelled guano masks from 24 Adélie penguin colonies. We also provide full metadata for images cropped from high-resolution imagery. These images include penguin colony crops from four different high-resolution optical satellites: GeoEye-1, QuickBird-2, Worldview-2 and Worldview-3. Depending on sensor, resolution for our images ranges from 2.4m/pixel (QuickBird-2) to 1.2m/pixel (Worldview-3) – the highest available on current commercial imagery.

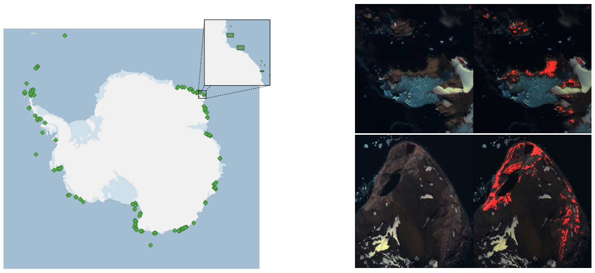
Adding our penguin colony polygons to medium-resolution Landsat-based masks from [18], we store the locations of 193 Adélie penguin colonies. With colony polygons in hand, we query an archive of 99653 high-resolution satellite images from the Antarctic coastline for images that encircle penguin colony shapes. We then crop each image to the smallest bounding box for each penguin colony, adding 100 pixels of padding on each side. For each cropped image generated this way, we calculate a Shannon entropy index, discarding crops that score 5 or lower. Following this automated data acquisition routine, we cropped 2044 images at locations shown in Fig. 2(a), heretofore referenced as the "weakly-annotated dataset". These 2044 images can be grouped into a video segmentation dataset [10, 23, 13, 33] consisting of image time series for each of the 193 penguin colonies in our dataset.

We then split the images from our 31 high-resolution masks into 18 training and 13 testing images. Similar to the weakly-annotated dataset, cropped images vary in size depending on the extent of the colony and the sensor resolution. Crops from high-resolution images for which we created segmentation masks are heretofore referenced as the "manually-labeled dataset".

In summary, we provide a dataset containing shapefiles for guano polygon masks and colony bounding boxes, and cropped images for manually-labelled and weakly-annotated penguin colonies. The weakly-annotated component of our dataset is easily expanded as our imagery archive grows. Though aircraft-based aerial imagery for related problems do exist (e.g. [1]), to the best of our knowledge, this is the first public dataset involving animal population estimation from high-resolution satellite imagery. More details can be found at: github.com/lynch-lab/CVPR19-CV4GC-WeaklyLabeling

4. Weakly-Supervised Learning for Penguin Colony Segmentation

As discussed in Section 3, only 0.8% (18 images) of our training data is hand-labelled, and there are 2044 penguin colony images with misaligned segmentation masks. There are two main challenges in this scenario: The first is that the image-level labels are unavoidably noisy. The images, although captured at the locations of known penguin colonies, might not contain visible penguin guano. The images could be covered in snow, shadows, or clouds, or were not captured in the breeding season when the penguin guano is visible. The second issue is that the pixel-level annotations are misaligned with the actual image contents due to georegistration errors or orthorectification artifacts.



(a) (b)

Figure 2. (a) Cropped image locations. Each square represents a colony bounding box (see inset) for which we found matching satellite imagery. To find matches, we query an archive of 99653 high-resolution imagery images obtained from 2002 to 2017. Our dataset harbors a total of 2044 satellite images, covering the vast majority of existing Adélie penguin colonies. (b) Two examples of hand-labelled guano mask (red overlay at right) on high-resolution imagery (left). Imagery copyright DigitalGlobe, Inc. 2019.

We propose a method to learn from those weakly-labeled data for image segmentation. We use two networks, C-Net and S-Net, to maximize learning from the hand-labelled data. C-Net is a classification network that learns to predict the image labels, e.g., whether an image contains ***any*** penguin guano and S-Net is a segmentation network that learns to segment the penguin guano areas. The main purpose of the C-Net is to filter out bad training examples from the weakly-annotated training set to better train the S-Net. Our framework is summarized in Fig. 3.

The C-Net first learns to classify images from the hand-annotated data. The training label for each image is binary: **0** implies the image is without any guano and **1** otherwise. Once the C-Net has been trained, we use it to assist the training of the S-Net. We train the S-Net on both the hand-labelled and the weakly-annotated data. For the hand-labelled data, the S-Net is trained to predict the segmentation mask from the input image. For each weakly-annotated image, we want to mitigate the risk of using bad training examples. Hence, we use the C-Net to classify all images that are weakly-labelled as containing guano and then remove all images that are classified as "no guano" from the training pool for the S-Net.

As minor georegistration errors or orthorectification artifacts create mismatches between annotation masks and input images, we do not use generated crops as groundtruth segmentation masks for S-Net. Instead, we train S-Net to recover the mean pixel values on weakly-annotated masks. Such a metric works as a proxy for fractional guano coverage in the images, which is more robust to imperfect georegistration. We essentially enforce an image-level statistic matching between predicted masks and weakly-annotated masks instead of minimizing pixel-wise differences.

Let I denote an input image, and $M(I)$ be the guano

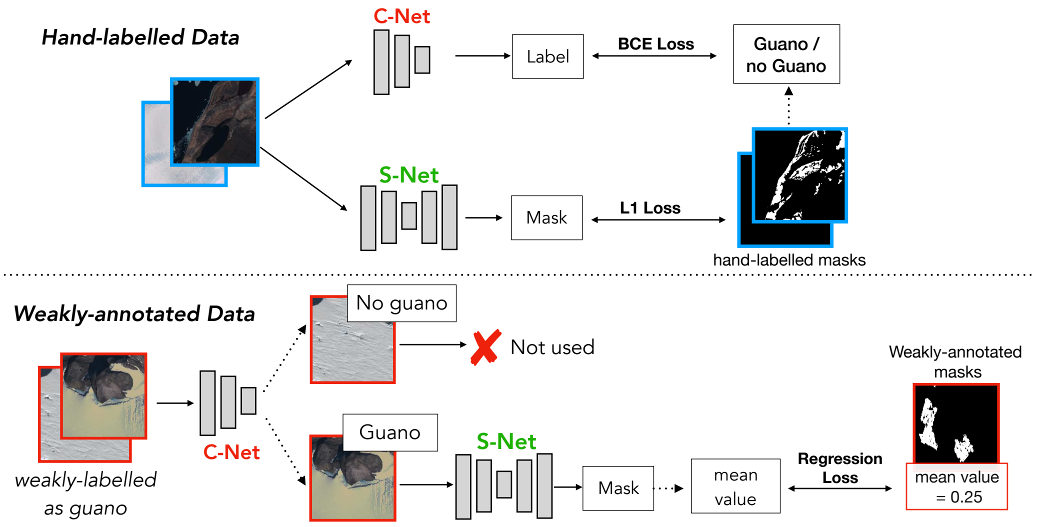


Figure 3. **Weakly-Supervised Learning Framework For Penguin Colony Segmentation.** Data with hand-labelled annotations are used to train both the C-Net and S-Net: The C-Net learns to predict the image label (if there is any penguin guano in the image) while the S-Net learns to segment the penguin guano areas. Once the C-Net is trained, it filters out images weakly-marked as containing guano but without visible guano due to snow, shadows, cloud, or poor timing relative to the breeding season. The S-Net learns from the weakly-annotated images to output the segmentation masks such that the mean activation of pixels in the predicted masks approximate the weakly-annotated masks. Imagery copyright DigitalGlobe, Inc. 2019.

mask of I. Let $S(I)$ denote the output of the S-Net for the input image I. Ideally, the output should be **1** for guano pixels and **0** otherwise. The objective of S-Net's training is to minimize a weighted combination of two losses:

$$\begin{aligned}\mathcal{L}_S(I) =& \lambda_{seg}\mathcal{H}(I)\left\|S(I) - M(I)\right\|_2 + \\ & \lambda_{reg}(1 - \mathcal{H}(I))\left\|\text{mean}(S(I)) - \text{mean}(M(I))\right\|_1\end{aligned}$$

where the value of $\mathcal{H}(I)$ is 1 if I is a hand-labelled image and 0 if I is a weakly annotated image. λ_{seg} and λ_{reg} control how much the S-Net should learn from the two losses respectively. We empirically set $(\lambda_{seg}, \lambda_{reg})$ to $(1, 5)$.

5. Experiments

We evaluate the performance of the C-Net and S-Net on the testing set of the Penguin Colony Dataset. The testing set contains 13 images of various sizes. To evaluate the performance of the C-Net, we crop the testing images into patches of size 256 × 256 with a step size of 64. The label for each image patch is obtained from the corresponding cropped mask. To train the S-Net, we crop each training image to small patches of size 384 × 384 with a step size of 192 to reduce I/O bottleneck issues arising due to the large sizes of images. From the original training set, we obtain 6055 hand-labelled and 100584 weakly-annotated training patches. To evaluate the performance of the S-Net, we compare the output of the S-Net to the hand-annotated guano masks of the Penguin Colony testing set.

We design the C-Net based on Resnet-18 [31] and the S-Net based on U-Net [24]. We use stochastic gradient descent with the Adam solver [11] to train our models.

Table 1. **C-Net classification results on the Penguin Colony Dataset.** Confusion matrices summarizing the results of C-Net on the image patches cropped from the Penguin Colony Dataset testing set (a) and on image patches cropped from the weakly-annotated set (b; W.A. Set). "Guano" patches contain penguin guano areas and "No Guano" patches contain only background. "*True Label*" means the patches are manually annotated. "*Weak Label*" means the labels are obtained via the possibly mismatched masks. An image patch is classified as positive if C-Net outputs a positive score and negative otherwise.

(a) Testing Set		*True Label*	
		Guano	No Guano
Pred.	Guano	858	469
	No Guano	265	13968

(b) W.A. Set		*Weak Label*	
		Guano	No Guano
Pred.	Guano	9597	2070
	No Guano	19446	69471

5.1. Penguin Guano Classification.

We first analyze the classification performance of the C-Net. We evaluate the C-Net on the testing set consisting of 15560 patches of sizes 256 × 256, which are cropped from 13 testing images with a step size of 64. Table 1(a) reports the confusion matrix summarizing the result of the C-Net on the testing image patches. The C-Net achieves a 0.65 precision and 0.76 recall on this set. For the weakly-annotated image patches shown in Table 1(b), C-Net classifies 19446 patches weakly-labelled as "Guano" to be non-guano. These patches then are not used for training the S-

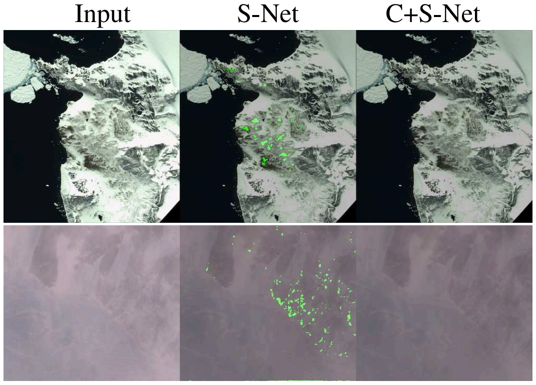


Figure 4. **The Effect of C-Net: Segmentation results of the S-Net trained on the whole dataset and on the dataset filtered by C-Net.** Input image (left column) covered by snow (top row) and fog (second row). Because the C-Net filters out noisy training examples for the S-Net, S-Net does not predict any guano pixels in either the snow or fog corrupted images. Imagery copyright DigitalGlobe, Inc. 2019.

Table 2. **Segmentation results of our models on the Penguin Colony Dataset.** We train our S-Net on different sets of training data and loss functions. “H” denotes hand-labelled data and “W” denotes weakly-annotated data. For “Seg. (H)”, we compute the segmentation loss on the hand-labelled data to train the network. For “Reg. (W)”, we compute the regression loss on the weakly-annotated data. “S.” only uses the segmentation loss while “SR.” uses the weighted combination loss of the segmentation and regression losses. “+” uses weakly-annotated data, and “C+S-Net” is our S-Net trained on the C-Net filtered data.

Method	Data	Loss Function	mIoU (%)
S-Net S.	H	Seg.(H)	42.3
S-Net S. +	H+W	Seg.(H) + Seg.(W)	37.7
S-Net SR. +	H+W	Seg.(H) + Reg.(W)	55.0
C+S-Net SR. +	H+W	Seg.(H) + Reg.(W)	**60.0**

Net. We show the performances of S-Net trained with and without these removed training patches in Section 5.2.

5.2. Penguin Guano Segmentation.

We evaluate our penguin colony segmentation network on the Penguin Colony Dataset. To obtain the output segmentation mask for an image, we first crop the image into patches of size 256×256 with a step size of 128. Each patch is input into the network to obtain a patch prediction mask. We obtain the final prediction mask for the input image by averaging all overlapped patch predictions at each pixel. We use mean Intersection-Over-Union (mIoU) to evaluate the segmentation masks.

Table 2 summarizes the results of our model. We compare our method with three baselines. All methods use as the backbone segmentation network the network with the same architecture as the S-Net. The first row shows the results of the S-Net trained on only the hand-labelled data, using the segmentation loss, denoted as “S-Net S.”. This model achieves 42.3 mIoU on the testing set of the Penguin Colony Dataset. A straightforward use of the weakly-annotated data is to train a segmentation network to output the segmentation masks, regardless of the mis-georegistration. This network is trained using the same segmentation loss function, but with more data, denoted as “S-Net S. +” in the second row. This model does not take into account the mis-georegistration issue of the guano polygons. Unsurprisingly, segmentation performance decreases from 42.3 to 37.7 mIoU score since the model is guided to output the guano pixels at the mismatched locations.

The third row shows the effect of our mean activation regression loss to learn from the misaligned guano masks. We train a S-Net on both the hand-annotated data and weakly-annotated data. For the weakly-annotated data, this S-Net is only constrained to output the segmentation masks that have the same average pixel values as the weakly-annotated guano masks. This model is denoted as “S-Net SR. +” in the third row of Table 2. This simple modification improves the mIoU by 30%, from 42.3 to 55 mIoU, compared to the model trained only on the hand-labelled training set.

We then evaluate the effect of the C-Net. We use C-Net to classify all 29043 training patches that are marked as containing guano, according to the weakly-annotated guano masks. We then remove image patches classified as “No Guano” from the training pool. As can be seen from Table 1, C-Net removes 19446 image patches, which are 67% of the whole set of positive training patches that are weakly-annotated. With less noisy training images, S-Net achieves better segmentation performance on the Penguin Colony Dataset where the mIoU improves from 55% to 60% mIoU score. Fig. 4 illustrates the effect of the C-Net. The figure shows that S-Net trained on the complete data is forced to predict the guano even on the snow covered areas while the model that trained on the filtered data, “C+S-Net”, does not predict any guano pixels in these cases.

Fig. 5 shows a qualitative comparison between our proposed method and the other three baselines discussed above. The three penguin colonies shown in the figure are (from top to bottom) Arthurson Ridge, Balaena Islands, and Cape Crozier. As can be seen, the model trained on only the hand-labelled data does not segment Arthurson Ridge and Cape Crozier correctly. The model trained without the C-Net picks up some non-guano pixels at Arthurson Ridge while missing some guano areas at Balaena Islands and Cape Crozier. The model trained on the data filtered by C-Net outputs cleaner and more complete segmentation masks.

The effect of using weakly-labelled data is shown best in Fig. 6. We compare the results of our model trained with and without the weakly-annotated data for images captured at the penguin site named Arthurson Ridge throughout mul-

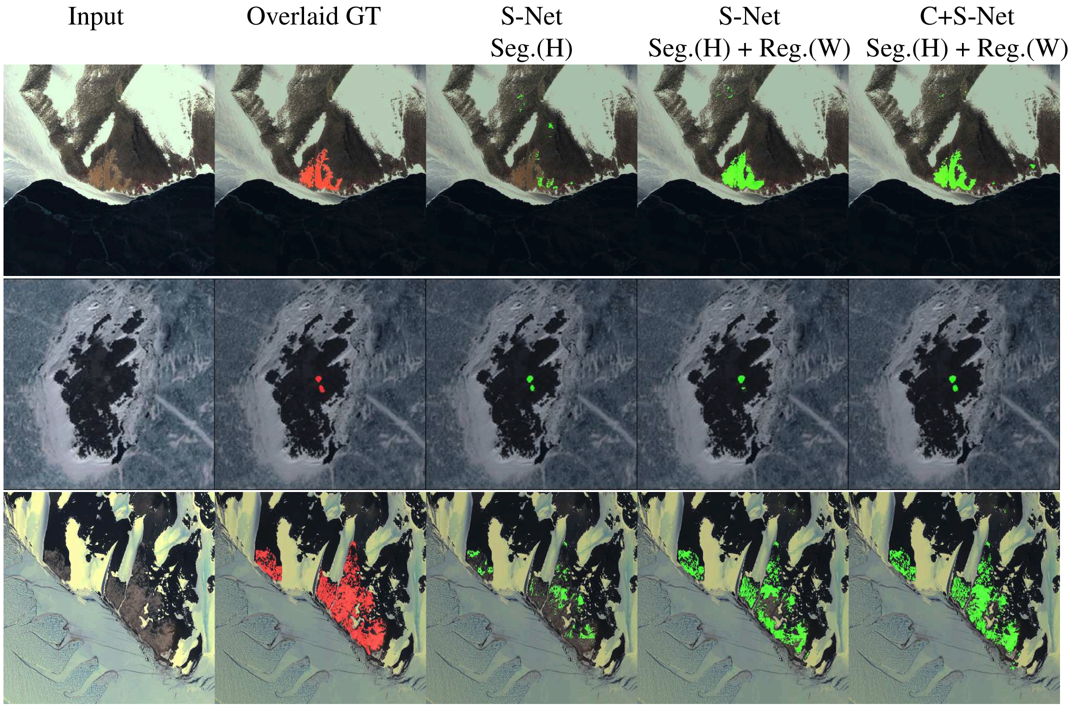


Figure 5. **Qualitative comparison between our method and other baseline methods on the Penguin Colony Dataset.** All methods use S-Net as the backbone segmentation network. From left to right: the input image, the input image overlaid by the ground truth penguin guano polygon, the results of S-Net trained only with the correctly-annotated set, the results of S-Net trained on the correctly-annotated set with the segmentation loss and on the weakly-annotated set with the regression loss. The last column is our proposed method using the C-Net to filter out bad training examples for training the S-Net on both correctly-annotated and weakly-annotated set. Imagery copyright DigitalGlobe, Inc. 2019.

tiple years. The weakly-annotated data improves significantly the generalizability of the network.

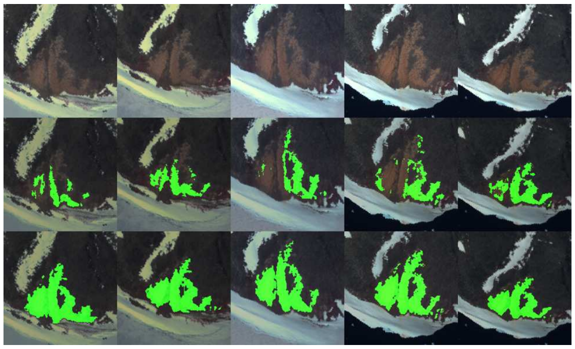


Figure 6. **Qualitative comparison of S-Net trained with and without weakly-labeled data.** The top row shows the input images, the middle row visualizes the results of S-Net trained only on the hand-labelled data, and the bottom row visualizes the results of S-Net trained on the hand-labelled data and weakly-annotated data. We used our trained C-Net to filter the training data before training this S-Net. Imagery copyright DigitalGlobe, Inc. 2019.

6. Conclusion

In this work, we present an approach to easily acquire weakly-annotated data on semantic segmentation datasets with georeferenced imagery. Our data acquisition routine circumvents limitations caused by unfavorable weather conditions and bad georegistration. We successfully weed out unsuitable training data using a CNN classifier, obtaining improved segmentation performance when using a subset of 9597 positive crops selected by our classifier, instead of training with all 29043 positive crops. Employing weakly-labeled data – curated by our CNN classifier – dramatically improves the accuracy of our segmentation output by 42%, increasing the mIoU from 42.3 to 60.0%. Apart from segmenting penguin guano, our framework is easily extensible to other georeferenced segmentation applications where we wish to track changes in features with predictable location.

Acknowledgements. We gratefully acknowledge assistance from the US National Science Foundation (Award No. NSF/ OPP-1255058, NSF/ ICER-1740595, NFS/ CNS-1718014), the National Geographic/Microsoft AI for Earth program, assistance from the Polar Geospatial Center, a gift from Adobe, the Partner University Fund, and the SUNY2020 Infrastructure Transportation Security Center. We would also like to acknowledge Stony Brook's Institute of Advanced Computational Science, the SeaWulf computing cluster and support staff, and Marcel Simon and Siyu Yang for helpful suggestions.

References

[1] Noaa fisheries steller sea lion population count | kaggle. www.kaggle.com/c/noaa-fisheries-steller- sea-lion-population-count/data. 3

[2] D. Ainley. *The Adélie penguin: bellwether of climate change*. Columbia University Press, 2002. 2

[3] E. Bjorgo. Refugee camp mapping using very high spatial resolution satellite sensor images. *Geocarto International*, 15(2):79–88, 2000. 1

[4] A. Borowicz, P. McDowall, C. Youngflesh, T. Sayre-McCord, G. Clucas, R. Herman, S. Forrest, M. Rider, M. Schwaller, T. Hart, et al. Multi-modal survey of adélie penguin mega-colonies reveals the danger islands as a seabird hotspot. *Scientific reports*, 8(1):3926, 2018. 2

[5] J. M. Buhmann. Weakly supervised structured output learning for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '12, 2012. 2

[6] M. A. Cimino, H. J. Lynch, V. S. Saba, and M. J. Oliver. Projected asymmetric response of adélie penguins to antarctic climate change. *Scientific reports*, 6:28785, 2016. 2

[7] J. Dai, K. He, and J. Sun. Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation. abs/1503.01640, 2015. 2

[8] J. Forcada and P. N. Trathan. Penguin responses to climate change in the southern ocean. *Global Change Biology*, 15(7):1618–1630, 2009. 2

[9] G. Humphries, R. Naveen, M. Schwaller, C. Che-Castaldo, P. McDowall, M. Schrimpf, and H. Lynch. Mapping application for penguin populations and projected dynamics (mapppd): data and tools for dynamic management and decision support. *Polar Record*, 53(2):160166, 2017. 2

[10] A. Khoreva, F. Galasso, M. Hein, and B. Schiele. Classifier based graph construction for video segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015)*, pages 951–960, Boston, MA USA, 2015. IEEE Computer Society. 3

[11] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *Proceedings of the International Conference on Learning Representations*, 2015. 4

[12] M. A. LaRue, H. Lynch, P. Lyver, K. Barton, D. Ainley, A. Pollard, W. Fraser, and G. Ballard. A method for estimating colony sizes of adélie penguins using remote sensing imagery. *Polar Biology*, 37(4):507–517, 2014. 2

[13] H. Le, V. Nguyen, C.-P. Yu, and D. Samaras. Geodesic distance histogram feature for video segmentation. In *Proceedings of the Asian Conference on Computer Vision*, nov 2016. 3

[14] H. Le, T. F. Y. Vicente, V. Nguyen, M. H. Nguyen, and D. Samaras. A+D Net: Training a shadow detector with adversarial shadow attenuation. In *Proceedings of the European Conference on Computer Vision*, 2018. 2

[15] H. Le, C.-P. Yu, G. Zelinsky, and D. Samaras. Colocalization with category-consistent cnn features and geodesic distance propagation. In *ICCV 2017 Workshop on CEFRL: Compact and Efficient Feature Representation and Learning in Computer Vision*, 2017. 2

[16] S. Liu, J. Feng, C. Domokos, H. Xu, J. Huang, Z. Hu, and S. Yan. Fashion parsing with weak color-category labels. *IEEE Transactions on Multimedia*, 16:253–265, 2014. 2

[17] Y. Liu, Z. Li, J. Tang, and H. Lu. Weakly-supervised dual clustering for image semantic segmentation. *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2075–2082, 2013. 2

[18] H. Lynch and M. LaRue. First global census of the adélie penguin. *The Auk*, 131(4):457–466, 2014. 2, 3

[19] H. J. Lynch, R. White, A. D. Black, and R. Naveen. Detection, differentiation, and abundance estimation of penguin species by high-resolution satellite imagery. *Polar Biology*, 35(6):963–968, February 2012. 1, 2

[20] K. Malkin, C. Robinson, L. Hou, and N. Jojic. Label super-resolution networks. In *International Conference on Learning Representations*, 2019. 2

[21] K.-K. Maninis, S. Caelles, J. Pont-Tuset, and L. V. Gool. Deep extreme cut: From extreme points to object segmentation. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 616–625, 2018. 2

[22] G. Papandreou, L.-C. Chen, K. Murphy, and A. L. Yuille. Weakly- and semi-supervised learning of a dcnn for semantic image segmentation. *arxiv*, 2015. 2

[23] F. Perazzi, A. Khoreva, R. Benenson, B. Schiele, and A.Sorkine-Hornung. Learning video object segmentation from static images. In *Computer Vision and Pattern Recognition*, 2017. 3

[24] O. Ronneberger, P.Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention*, volume 9351 of *LNCS*, pages 234–241, 2015. 4

[25] O. Russakovsky, A. L. Bearman, V. Ferrari, and L. Fei-Fei. What's the point: Semantic segmentation with point supervision. In *ECCV*, 2016. 2

[26] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb. Learning from simulated and unsupervised images through adversarial training. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016. 2

[27] M. Tang, A. Djelouah, F. Perazzi, Y. Boykov, and C. Schroers. Normalized cut loss for weakly-supervised cnn segmentation. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1818–1827, 2018. 2

[28] A. Vezhnevets, V. Ferrari, and J. M. Buhmann. Weakly supervised semantic segmentation with a multi-image model. *2011 International Conference on Computer Vision*, pages 643–650, 2011. 2

[29] T. F. Y. Vicente, L. Hou, C.-P. Yu, M. Hoai, and D. Samaras. Large-scale training of shadow detectors with noisily-annotated shadow examples. In *Proceedings of the European Conference on Computer Vision*, 2016. 2

[30] C. Witharana and H. Lynch. An object-based image analysis approach for detecting penguin guano in very high spatial resolution satellite images. *Remote Sensing*, 8(5):375, 2016. 2

[31] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He. Aggregated residual transformations for deep neural networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 4

[32] J. Xu, A. G. Schwing, and R. Urtasun. Learning to segment under various forms of weak supervision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 2015. 2

[33] C.-P. Yu, H. Le, G. Zelinsky, and D. Samaras. Efficient video segmentation using parametric graph partitioning. In *ICCV*, 2015. 3