6.s898 - Climate Change Seminar
Amazon Land-cover classification from high resolution satellite imagery

Karan Bhuwalka (bhuwalka@mit.edu)
Technology and Policy Program, MIT
(Dated: December 11, 2019)

In this paper, I tackle a multi-label classification problem for the Amazonian forest. I aim at labeling satellite chips with feature tags corresponding to land cover and land use after clearing for atmospheric conditions. Using a recently disclosed labeled data set from Planet consisting of 40,000 256 x 256 x 4 satellite chips, I use several convolutional neural networks to generate classifications for each chip. I attempt to explore the sensitivity and robustness of the model with regard to a number of architecture parameters, using ResNet50 as a basis. I reach an f2 score of around 0.93 on the Planet data set and show the potential of the model for easily available data sets with lower resolution.

I. MOTIVATION

The original motivation of this project came when talking to a colleague of mine working in Prof. Dava Newman's lab. He described a project where they want to use computer vision techniques to generate hypothetical future satellite images under different climate change scenarios. I thought it would be an interesting class project to learn how to procure, process and analyze satellite imagery and study deforestation. As we've discussed in class, analysis of satellite imagery is a major way modern computing can assist climate change efforts. Classify forest cover loss due to mining would also be valuable for my own research interests on the societal impacts of mining. This project covers my attempts to train a model that classifies land cover in the Amazon.

II. CONTEXT AND PROBLEM STATEMENT

As deforestation now turns into unprecedented scales, the motives underlying the decline in forest resources need to be better defined and quantified [1]. As a matter of fact, growing demand for food resource drive the development of new agriculture lands, whether for crops or pasture in area previously covered by primary or secondary forests. In order to inform pro-forest policy-making, local governments and international organizations have long tried to quantify the value associated with deforestation, by assessing both the magnitude and different uses of deforested areas, with limited results[2].

In practice, with advances in satellite imagery, detection of deforestation has very recently experienced significant improvements. The Real Time System for Detection of Deforestation (DETER), which has been credited for reducing the deforestation rate in Brazil by almost 80 percent since 2004 by raising awareness of the environmental police to large-scale forest clearing[3]. However, current tracking efforts for deforestation mostly rely son coarse-resolution imagery from Landsat (30 meter pixels) or MODIS (250 meter pixels). The challenges faced by these methods are the limited effectiveness in detecting small-scale deforestation or differentiating between human causes of forest loss and natural causes. On the other hand, Planet, a provider of space imagery satellites has a labelled data set of land surfaces working at a 3-5 meter resolution.

This paper aims at providing possible methods to answer this quantification need. It focuses on the world's region most affected by deforestation, the Amazonian Basin[4]. Building on expanding remote sensing imagery data bases, it aims at providing a general algorithmic tool to detect and quantify, in given images of the same rain forest region over different periods of time, the change in forest coverage in the Amazon and as well as the most typical uses of deforested areas over time. I propose to leverage modern deep learning techniques to identify activities happening within the images. The problem is treated as a multi-label classification problem, and I aim to label satellite image chips with one to 16 labels indicating land cover modes and land use patterns.

III. LITERATURE REVIEW

The literature on deforestation has been quickly growing over the recent years. There is a large literature corpus covering the main drivers of deforestation and the different space imagery and remote sensing methods that could be applied in order to detect the delimitation of forests as well as agricultural activities supplanting previously forested areas.

First of all, Brazil appears as both a natural case study for my project but also as a well-document area well suited to a remote sensing analysis of deforestation and land use patterns. Parente et al.'[5] as well as Muller et al. [6] proposes a Random Forest model in order to identify zones, using features such as the vegetation index on 500 manually classified data points in predefined areas. What is more, [7] further elaborates on the effectiveness of random forests and convolutional neural networks (CNNs), and shows that these two latter can yield similar results for classification and propose various methods aiming at classifying land use and different possible crops.

Also, the literature on image pre-processing provides a broad overview of the different methods available to facilitate the detection and quantification of land use features that could otherwise be obstructed by adversarial atmospheric features. He et al.[8] proposes a single image haze removal method leveraging dark channel priors, separating a "clear" image from its "hazy" components. Additionally, Enamoto et al.[9] recognizes the impac tof clouds on the degradation of accuracy regarding the detection of land use patterns using spae imagery, and details a cloud removal methodology based on multispectral conditional generative adversarial nets (GANs).

Lastly, a growing literature is now emerging regarding concrete algorithms for the detection of deforestation and land use. Cohen et al. demonstrated as early as 1998 that Landsat imagery can be used to map forest clear cut sin the Pacific Northwest[10]. Popatov et al.[11] and Hansen et al.[12] combined MODIS and Landsat data to estimate forest cover change in boreal forests and the Congo Basin, further expanding the use of remote sensing data to new geographies. More recently, the urgency of tackling deforestation has also led to the Brazilian authorities to establish their own real-time system monitoring forest clearing, called DETER and PRODES[11].

More specifically, the recent years have seen a sharp increase in the machine learning methods to analyze satellite data. Mnih and Hinton[13] used large-scale neural networks to detect roads in high-resolution aerial images. In addition, Kluckner et al.[14] applied covariance descriptors to allow for the multiclass classification of aerial images.

IV. DATASET AND FEATURES

To train my model, I use 3.5m resolution satellite data from Planet's[15] full frame analytic scenes using 4-band satellites in sun-synchronous orbit (SSO) and International Space Station (ISS) orbit. This data is now available to MIT students and staff. The training, validation and testing data consists of labelled 40479 256x256 resolution (957m x 957m) chips that Planet released in 2017 for a Kaggle challenge[16]. All of the scenes come from the Amazon basin which includes Brazil, Peru, Uruguay, Colombia, Venezuela, Guyana, Bolivia, and Ecuador. Each chip has been manually labelled by the Planet team (alongwith crowd-sourced labour) into 17 categories.

The 17 labels consist of 5 weather labels and 12 land-use categories. The histogram Fig 1 shows the prevalance of each of the labels in the dataset. Each image can have multiple labels. Figure 2 show a few samples images

I will also test my model on lower resolution Google Earth data which is based on the Landsat8 satellite.

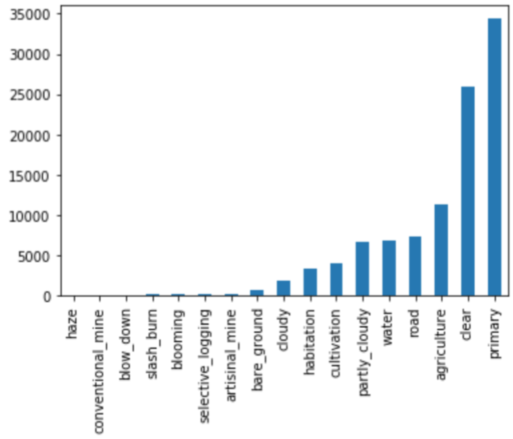

FIG. 1: Prevalance of labels in the dataset

FIG. 2: Sample images with labels

V. METHODS

A. Pre-processing

The first step was to 'de-haze' my images so that the model could see the images more clearly. I apply this to all images, even the ones not explicitly classified as haze because I find in the data that images like in Figure 3 can be labelled as clear but still look hazy. To do this I applied the algorithm from 'Single Image Haze Removal Using Weak Dark Channel Prior' as described by Hsieh et al [17]. This method uses the dark channel prior to estimate the atmospheric light in an image, which can then be used for removing haze. An example of this applied to my data is shown in Figure 3, where the left hand image is the original data and the right hand side is de-hazed. After dehazing, I manually delete the 'haze' label.

B. Training

I split the data 70:20:10 into training, validation and testing. The neural network architecture I picked was the

FIG. 3: Haze Removal Using Weak Dark Channel Prior

Resnet50 architecture, as a lot of the top results from the Kaggle challenge seemed to get good performance from it. The network was trained using the fastai package. To determine a good learning rate, at each stage of training, I generated curves of Loss vs LR to find small learning rates such that log-loss goes down upon training. I conducted training in two batches: first I trained on a lower 128x128 resolution version of the images and then trained the same number of cycles on a full 256x256 image. I did this I want my model to perform well on both high and low resolution images. The optimizer I use is 'Adams Optimizer' which updates the learning rate based on the sparseness of the parameter. I provide the initital learning rate as a hyperparameter to the optimizer.

I first trained the last layer at a Learning Rate of 0.1 for 5 epochs, using 128x128 resolution data. I then unfroze the network and trained the entire network with a LR ranging from 10^{-5} for the last layers to 0.1 for the first layer for 5 epochs. I then froze the newtork and trained the last layer for 5 epochs with 256x256 data and an LR of 0.01. my final training round was for the unfrozen network with a rate going from 10^{-5} to 0.01 over the network for 5 epochs. Figure 4 shows the training and validation losses over time.

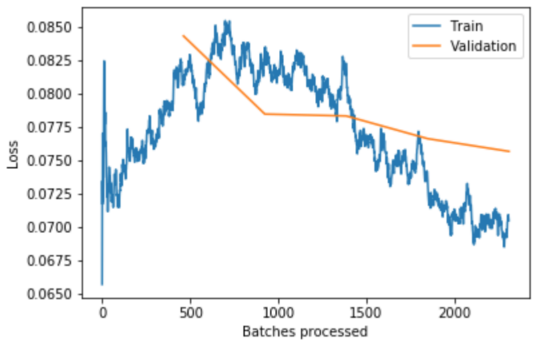

FIG. 4: Losses for Training and Validation

C. Evaluation metrics

Using pure accuracy (percentage of correct positive and negative guesses) as a metric of evaluation for my model is not ideal. This is because most images have primary forest and simply guessing 'primary' for every image would give us an 'accuracy' of 92.8. Similarly for low presence labels, guessing their absence would give us high accuracy.

I use the F2 score as a metric to evaluate the model. F-beta score is defined by the formula: $F_\beta = \frac{(1+\beta^2)*P*R}{(\beta^2*P+R)}$ where beta is a hyper parameter, P is precision and R is recall. Precision is defined as the ratio of True Positives and True Positives +False Positives. Recall is defined as the ratio of True Positives and True Positive + False Negative. Recall measures how many of the labels in an image were identified, while precision measures how many guesses were not present in the image.

I use an F2 score with β=2 so that I value recall over precision. I think it is more important to identify everything in an image rather than be selective or precise about my labels. This is because I want to ensure I identify all the labels in a multilabel classification, even if it comes at the cost of making a few incorrect guesses. An example of when this is important is the case of the forest fires, where I want to detect fires early even when the model thinks there is a relatively low probability of a fire actually existing.

I also try to estimate the optimal probability threshold for making a prediciton. Depending on the evaluation metric, I have different thresholds. I get the highest F2 score on the validation set of 0.93007 when the threshold is kept at P=0.19 . Meanwhile, for an F1 score which is symmetric about precision and recall I get a score of 0.91 at a higher threshold of P=0.42. For basic accuracy measurement, the threshold is even higher at 0.54 for an accuracy of 0.97. I can see this in figure 5

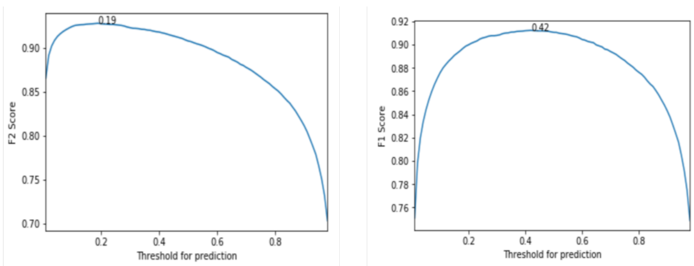

FIG. 5: Model Performance at different prediction thresholds. (Left) F2 Score (Right) F1 Score

VI. RESULTS

A. Overall Model Performance

I get an F2 score of 0.9280 on the test set. The F1 score is 0.9121 and the accuracy is 0.9692. On using Test Time Augmentation, where predictions are made by averaging over many random transformations (rotations, flips etc.) on the test image, the F2 score goes up to 0.9282.

The table I describes results of F1,F2 and accuracies for each label. I can see that in general, decreasing samples reduces performance.

Class	F2	F1	Accuracy	Samples
primary	0.990294	0.983212	0.967184	34338
clear	0.976552	0.968005	0.952194	25990
agriculture	0.896755	0.849718	0.907765	11300
road	0.864270	0.838291	0.937070	7375
water	0.822313	0.796467	0.932748	6795
partly_cloudy	0.941860	0.915523	0.969210	6667
cultivation	0.705156	0.619661	0.925996	4106
habitation	0.780502	0.715935	0.950034	3320
cloudy	0.835594	0.806723	0.981769	1892
bare_ground	0.435520	0.288136	0.978393	775
artisinal_mine	0.887574	0.902256	0.998244	312
selective_logging	0.394737	0.370370	0.993788	311
blooming	0.246711	0.057971	0.991897	304
slash_burn	0.273632	0.272727	0.994868	188
blow_down	0.259740	0.260870	0.997839	92
conventional_mine	0.523256	0.533333	0.997974	90

TABLE I: Results for each label

B. Case study on a Mine in Ecuador

An example of predictions made by my model can be seen in the case study of the Mirador Mine in Ecuador. This is a newly opened copper mine, which has caused huge controversy due to displacement of indigenous people. As shown in Figure 6, the model can catch change of forest into mine and then the addition of roads and water source to the mine.

Applying this land-use classification to a nearby larger 400km^2 area in 2018 and 2019, I can see how land cover changes. I estimate the area of a particular land cover type based on the number of 256x256 tiles that are classified to have that label. This is illustrated in the chart 7, where I can see the reduction in primary forest area and the increase in cultivation. Interestingly, the increase in cultivation in 2019 is preceded by logging in 2018, suggesting that this could be a good method for early detection of deforestation.

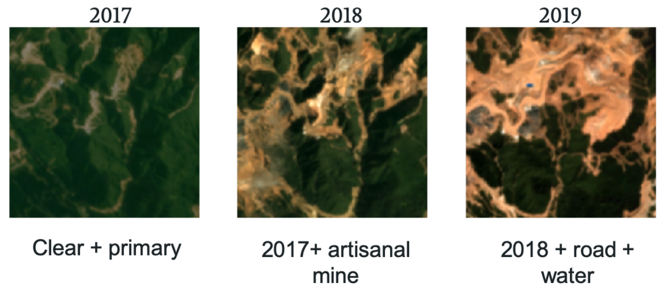

FIG. 6: Prediction for Mirador Mine from 2017-2019

C. Identifying Sources of Error

The first major source of error in my estimations is the presence of clouds. As you could see from Figure 7, different images have different amount of cloud cover and this makes it hard to compare images directly.

Secondly, while I try to train my image on low resolution versions of the image as well, it does not perform as well on Google Earth imagery and I are potentially restricted to using high resolution imagery from planet

A third issue is that the satellite doesnt go over the exact same area on each pass. This means that it is difficult to get two scenes with the exact same boundaries. However, if I get an area large enough to stitch multiple scenes, this error should reduce

Then there is the issue of errors in the model itself. Looking at the confusion matrix in Figure 8, I can see that the model does not do well on scenes where there is no forest. When there is no forest, it predicts the existence of forest 42% of the time This is because most of my training set has forest in it, plus I do not value precision very much. With more training data of non-forest scenes, my model should get better. Also from table I I know that I are not very successful at predicting rare images. This could become an issue in specific applications such as detection of relatively rare events like forest fires. For most of the land cover classes I care about however, my model predicts between 86%-95% of the labels accurately.

Finally, it is difficult to generalise these results outside the Amazon. On trying my model on images of the Tundra, my performance was not good.

VII. FUTURE WORK, CONCLUSION AND LEARNINGS

A major scope of future work is cloud removal and image reconstruction. This would be very useful in daily land-use change detection. I can also use probabilistic

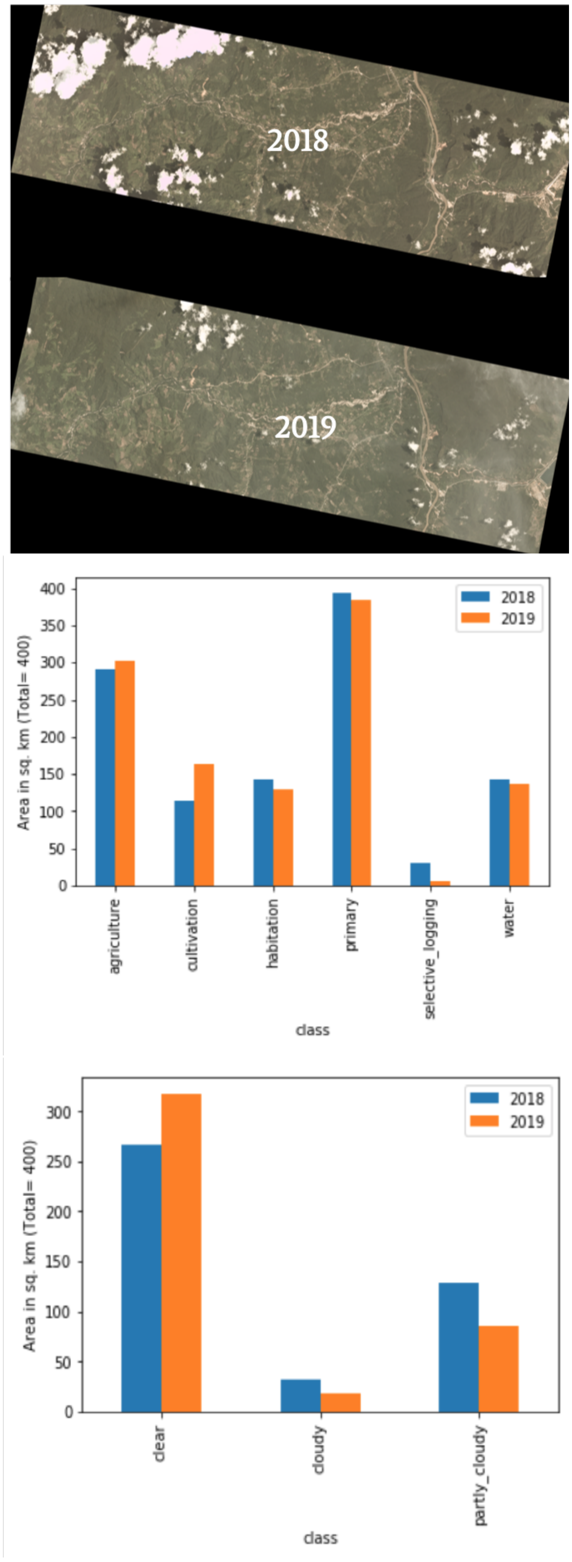

FIG. 7: Prediction of land cover change from 2018-2019

models that use information from nearby grids to inform the prediction of the current scene.

Advanced machine learning algorithms can also allow

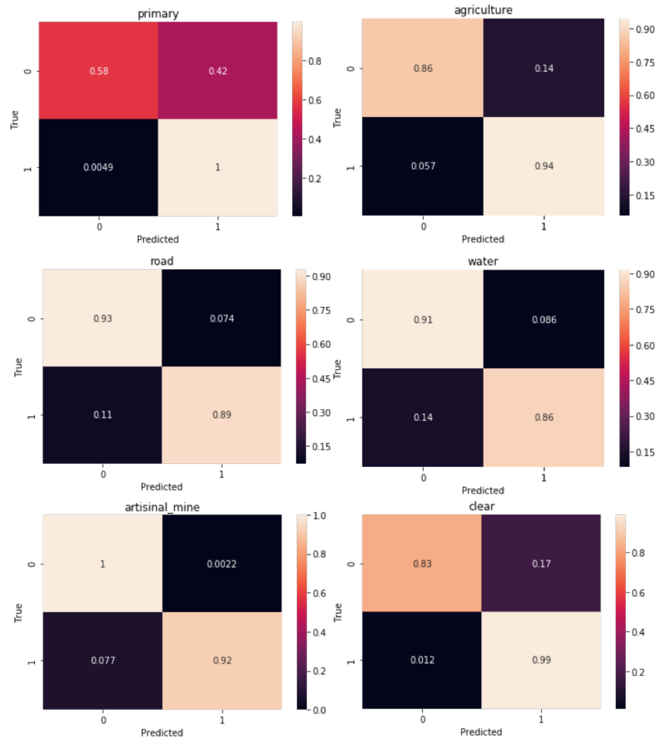

FIG. 8: Confusion matrices for selected labels

us to transform satellite images and create 'fake' images under future climate change scenarios for communication purposes. An example of this kind of work is GANPaint[18] [19] which allows us to change trees based on the season as shown in Figure 9

FIG. 9: Changing seasons with GANPaint

Using high-resolution satellite imagery has immense potential for many different applications. It can help identify sources of deforestation and make policy recommendation. An increased resolution can help local policymakers. It can also be used to identify illegal mining and ensure that supply of materials is done sustainably. MIT's new Planet license gives us all the opportunity to use this as a dataset in environmental research and communication. Being able to use visual imagery to display planetary degradation is a vital visceral tool.

I myself had no experience with using satellite imagery or neural networks before this class. It is something I've picked up and will use in future research on the environmental impacts of mining. I would look forward to talk to others who are looking to use this method in their work and help them get started.

---

[1] Jean Ometto, Ana Aguiar, and Luiz Martinelli. Amazon deforestation in brazil: Effects, drivers and challenges. *Carbon Management*, 2:575–585, 04 2014.

[2] Jeffrey Hoelle. Quantifying cultural values associated with deforestation in the brazilian amazon. *Journal of Land Use Science*, 13(1-2):166–181, 2018.

[3] Gabriel Popkin. Satellite alerts track deforestation in real time. *Nature*, 530:392–393, 2016.

[4] Herton Escobar. Amazon fires clearly linked to deforestation, scientists say. *Science*, 365(6456):853–853, 2019.

[5] Leandro Parente, Laerte Ferreira, Adriano Faria, Sérgio Nogueira, Fernando Araújo, Lana Teixeira, and Stephen Hagen. Monitoring the brazilian pasturelands: A new mapping approach based on the landsat 8 spectral and temporal domains. *International Journal of Applied Earth Observation and Geoinformation*, 2017.

[6] Hannes Müller, Philippe Rufin, Patrick Griffiths, Auberto José Barros Siqueira, and Patrick Hostert. Mining dense Landsat time series for separating cropland and pasture in a heterogeneous Brazilian savanna landscape. *Remote Sensing of Environment*, 2015.

[7] Edwin Raczko and Bogdan Zagajewski. Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral apex images. *European Journal of Remote Sensing*, 50:144–154, 2011.

[8] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2341–2353, Dec 2011.

[9] Kenji Enomoto, Ken Sakurada, Weimin Wang, Hiroshi Fukui, Masashi Matsuoka, Ryosuke Nakamura, and Nobuo Kawaguchi. Filmy cloud removal on satellite imagery with multispectral conditional generative adversarial nets. *CoRR*, abs/1710.04835, 2017.

[10] Maria; Gray John; Helmer Eileen; Anderson Karen Cohen, Warren B.; Fiorella. An efficient and accurate method for mapping forest clear cuts in the pacific northwest using landsat imagery. *Photogrammetric Engineering and Remote Sensing*, pages 293–300, 1998.

[11] Gabriel Popkin. Fusing landsat and modis data to better estimate boreal forest loss. page 3708–3719, 2008.

[12] Thomas Hilker, Michael A Wulder, Nicholas C Coops, Julia Linke, Greg McDermid, Jeffrey G Masek, Feng Gao, and Joanne C White. A new data fusion model for high spatial-and temporal-resolution mapping of forest disturbance based on landsat and modis. *Remote Sensing of Environment*, 113(8):1613–1627, 2009.

[13] Volodymyr Mnih and Geoffrey E Hinton. Learning to detect roads in high-resolution aerial images. In *European Conference on Computer Vision*, pages 210–223. Springer, 2010.

[14] Stefan Kluckner, Thomas Mauthner, Peter M Roth, and Horst Bischof. Semantic classification in aerial imagery by integrating appearance and height information. In *Asian Conference on Computer Vision*, pages 477–488. Springer, 2009.

[15] Planet explorer. `https://www.planet.com/explorer/`.

[16] Planet data for kaggle challenge. `https://www.kaggle.com/c/planet-understanding-the-amazon-from-space/data`.

[17] Cheng Hsiung Hsieh, Qiangfu Zhao, and Wen Chang Cheng. Single Image Haze Removal Using Weak Dark Channel Prior. In *2018 9th International Conference on Awareness Science and Technology, iCAST 2018*, 2018.

[18] Gan paint. `http://ganpaint.io/#paper`.

[19] David Bau, Hendrik Strobelt, William Peebles, Jonas Wulff, Bolei Zhou, Jun Yan Zhu, and Antonio Torralba. Semantic photo manipulation with a generative image prior. *ACM Transactions on Graphics*, 2019.