SEGMENTATION OF SINGLE STANDING DEAD TREES IN HIGH-RESOLUTION AERIAL IMAGERY WITH GENERATIVE ADVERSARIAL NETWORK-BASED SHAPE PRIORS

P. Polewski[1], J. Shelton[1], W. Yao[1,*], M. Heurich[2]

[1] Dept. of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China
- (przemyslaw.polewski, jacquelyn.ann.shelton)@gmail.com, wei.hn.yao@polyu.edu.hk
[2] Dept. for Conservation and Research, Bavarian Forest National Park, Grafenau, Germany
- marco.heurich@npv-bw.bayern.de

KEY WORDS: GAN, CNN, active contour, color infrared, forest health monitoring

ABSTRACT:

The use of multispectral imagery for monitoring biodiversity in ecosystems is becoming widespread. A key parameter of forest ecosystems is the distribution of dead wood. This work addresses the segmentation of individual dead tree crowns in nadir-view aerial infrared imagery. While dead vegetation produces a distinct spectral response in the near infrared band, separating adjacent trees within large swaths of dead stands remains a challenge. We tackle this problem by casting the segmentation task within the active contour framework, a mathematical formulation combining learned models of the object's shape and appearance as prior information. We explore the use of a deep convolutional generative adversarial network (DCGAN) in the role of the shape model, replacing the original linear mixture-of-eigenshapes formulation. Also, we rely on probabilities obtained from a deep fully convolutional network (FCN) as the appearance prior. Experiments conducted on manually labeled reference polygons show that the DCGAN is able to learn a low-dimensional manifold of tree crown shapes, outperforming the eigenshape model with respect to the similarity of the reproduced and referenced shapes on about 45 % of the test samples. The DCGAN is successful mostly for less convex shapes, whereas the baseline remains superior for more regular tree crown polygons.

1. INTRODUCTION

From an ecological perspective, monitoring the state and quantity of coarse woody debris (CWD) is a crucial task due to its role in forest biodiversity, nutrient cycles, and as carbon sequestration (Harmon et al., 1986). It is well known that dead vegetation produces a distinct reflectance signature in the infrared spectral band (Jensen, 2006), therefore remote sensing based on passive optical sensors has been widely used for detecting diseased and dead trees (Wang et al., 2007; Vogelmann, 1990; Heurich et al., 2010; Polewski et al., 2016). Advances in optical sensor technology have capacitated the widespread use of high-resolution aerial images in large-scale forest inventories. This opens up new possibilities for mapping dead vegetation with unprecedented precision and spatial coverage. Although the infrared spectral band is useful for detecting dead vegetation, there are still a number of challenges associated with extracting individual dead trees. First, there may be other objects within the scene which possess a similar reflectance signature (e.g. open ground patches, roads etc.). Also, the centimeter-resolution aerial imagery reveals much more complexity in the shape of dead tree crowns than was possible with the previous generations of sensors. This calls for appropriately complex shape models, capable of representing the entire range of the tree crown variability. Finally, the interactions of adjacent dead tree crowns can lead to the formation of complex aggregates that are difficult to separate into individual trees.

Recently, convolutional neural networks (CNNs) have become the vanilla standard for many classical computer vision tasks. Dense semantic segmentation of raster imagery is particularly well handled by fully convolutional encoder-decoder network architectures such as the U-Net (Ronneberger et al., 2015). Another area where CNNs have achieved spectacular success is generative modeling of image distributions pertaining to a particular domain. Generative adversarial networks (GANs) (Goodfellow et al., 2014) are able to learn highly complex mappings from a low dimensional latent space Z to the image manifold X, such that sampling from Z induces a corresponding set of samples from X. On the other hand, segmenting the scene into individual objects remains a difficult and challenging problem and an active area of research within the neural network and computer vision community (Arnab and Torr, 2017; He et al., 2017). Current state-of-the-art CNN based approaches suffer from coarseness of feature maps as well as limited information contained in the candidate object regions of interest, resulting in degraded performance for small and multi-scale object localization (Zhao et al., 2018).

We propose to combine the strengths of GANs and fully convolutional networks within the energy minimization framework of active contour segmentation (Cremers et al., 2007). This mathematical formalism describes the evolution of a segmented object's contour within an image, partitioning the image into regions inside and outside of the target object. The segmentation process may be endowed with prior information. Specifically, we utilize the probabilistic formulation of Cremers and Rousson (2007), which admits explicit priors for the object shape and appearance. However, we replace the original linear eigenshape model of Tsai et al. (2003) with a deep convolutional GAN (Radford et al., 2016). Also, the target class posterior probability output from a U-Net (Ronneberger et al., 2015) is used in the role of the appearance prior instead of the kernel density estimator models of intensity used by the authors earlier. This is a modification of our prior work (Polewski et al.,

---

2015), where simpler, non CNN-based priors were used. We tested our approach on 200 polygons manually marked within high-resolution color infrared (CIR) imagery from the Bavarian Forest National Park. We evaluated the segmentation approach enriched with CNN-based priors against the baseline method with respect to both per-pixel similarity of the reference and segmented polygons and more abstract shape similarity measures.

The rest of this paper is organized as follows. In Section 2, we review previous work with the conceptually closest approaches regarding both dead/diseased tree detection from multispectral imagery and combining active contour segmentation with deep learning approaches. Section 3 explains the general framework of active contour segmentation and in particular the mechanisms of incorporating prior information. The next section deals with the architecture of the applied GAN shape prior. We describe the computational experiment, source data and evaluation metrics in Section 5, and in Section 6 we describe the results and discuss them. In the final section we state the key findings and conclusions of our work.

2. RELATED WORK

Several authors attempt segmentation and classification of individual dead or diseased trees from aerial infrared imagery. Bhattarai et al. (2012) first apply a generic individual tree crown segmentation, and subsequently classify each tree as either dead or living based on multispectral features. More recently, Safonova et al. (2019) used a CNN to assess the vitality of trees from aerial RGB images. They used rectangular patches as the data unit for classification. Näsi et al. (2018) report using hyperspectral imagery for identifying dead and diseased trees. Their comparison between data acquired by means of an unmanned aerial vehicle (UAV) versus aircraft-mounted sensors revealed that the superior UAV ground sampling distance of 10 cm yields significant improvements in detection accuracy. This bolsters our working hypothesis that the increased image resolution translates to more information content relevant to the segmentation task. The listed approaches share the characteristic of splitting the dead tree segmentation problem into a generic tree crown delineation step, followed by a classification phase. We are not aware of any competing approaches which attempt to explicitly model the dead tree shape and make use of it during the search for dead trees, except our own prior work.

There also exists prior work on the topic of combining convolutional neural networks with active contour segmentation. Marcos et al. (2018) proposed a framework utilizing CNNs for learning the geometric prior parameters of an active contour model in the context of single instance segmentation of urban scenes containing buildings. They showed how the CNN training task can be cast as a structured learning problem, enabling end-to-end training.

This work draws upon some ideas presented in (Wu et al., 2017), where a GAN-like model was utilized to generate 3D objects based on silhouette and surface normal information, but decoupled from the object's texture/appearance. Also, in some sense our work concerns the problem of inverting GAN models, i.e. finding a latent variable vector which leads to the generation of a given input object (e.g. image). This is also an ongoing topic within the neural network community (Creswell and Bharath, 2019).

3. ACTIVE CONTOUR SEGMENTATION

3.1 General setting

In the setting of image segmentation, let $\Omega \subset R^2$ be the image plane, $I : \Omega \to R^d$ a vector-valued image, and C an evolving contour in the image I. We wish to find a contour C which partitions the image 'optimally' into two disjoint regions Ω_1 and Ω_2, such that the former represents the 'foreground', or part of the image located within C, and the latter represents the background. The notion of optimality may be expressed in a probabilistic fashion using the Bayesian rule:

$$\mathcal{P}(\Omega_1, \Omega_2 | I) \propto \mathcal{P}(I | \Omega_1, \Omega_2) P(\Omega_1, \Omega_2) \tag{1}$$

Furthermore, the contour C uniquely identifies the partition Ω_1, Ω_2, therefore the (log) probability of the partition given the image data decomposes into a shape prior term and a data likelihood term:

$$
\begin{aligned}
E(C) &= -log(\mathcal{P}(C|I)) = -log(\mathcal{P}(I|C)) - log(\mathcal{P}(C)) \\
&= E_{img} + E_{shp} \\
E_{img} &= -log(\mathcal{P}(I|C)), E_{shp} = -log(\mathcal{P}(C))
\end{aligned} \tag{2}
$$

Assuming that region labellings are uncorrelated, i.e. $\mathcal{P}(I|\Omega_1, \Omega_2) = \mathcal{P}(I|\Omega_1)\mathcal{P}(I|\Omega_2)$, and also that image values inside a region are realizations of independent and identically distributed random variables (Cremers et al., 2007), the image term becomes:

$$
\begin{aligned}
E_{img}(C) &= -\int_{\Omega_1} \log f_1(I(x))dx - \int_{\Omega_2} \log f_2(I(x))dx \\
&= -\int_{\Omega} \mathbb{1}_C[x] \log f_1(I(x)) + \\
&\quad (1 - \mathbb{1}_C[x]) \log f_2(I(x))dx
\end{aligned} \tag{3}
$$

In the above f_1 and f_2 denote the probability density functions of the image values inside and outside the contour C (corresponding respectively to the regions Ω_1, Ω_2), whereas $\mathbb{1}_C(z)$ is the indicator function for the set Ω_1 (i.e. interior of C). We can further represent the generative pixel probability f_i for region Ω_i in terms of the (discriminative) region label posterior:

$$f_i(I(x)) = \mathcal{P}(I(x)|x \in \Omega_i) \propto \mathcal{P}(x \in \Omega_i|I(x))\mathcal{P}(x \in \Omega_i) \tag{4}$$

In the above, the probability of observing an image value $I(x)$ has been dropped as independent from the contour. Assuming that the probability of an image element x belonging to the foreground or background is not dependent of the position of x within the image, we can drop the term $\mathcal{P}(x \in \Omega_i)$ and utilize the class posterior within the image energy term (Polewski et al., 2015), resulting in the familiar binary cross-entropy:

$$
\begin{aligned}
E_{img}(C) &= -\int_{\Omega} \mathbb{1}_C[x] \log \mathcal{P}(x \in \Omega_1|I(x)) + \\
&\quad (1 - \mathbb{1}_C[x]) \log[1 - \mathcal{P}(x \in \Omega_1|I(x))]dx
\end{aligned} \tag{5}
$$

3.2 Level-set segmentation

In the level-set formulation, the contour C itself is not explicitly evolved. Rather, it is assumed that the contour is implicitly represented as the 0-th level set of an embedding function ϕ :
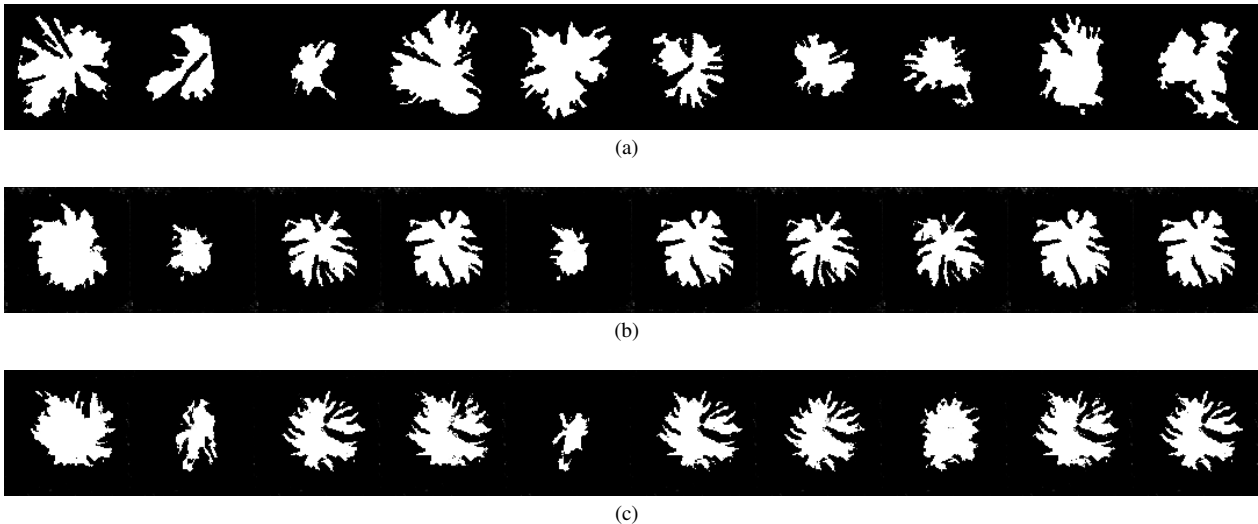
Figure 1. (a) sample training shapes for GAN model, (b)-(c) shape samples generated by DCGAN after (b) 1500 and (c) 3500 iterations.

$\Omega \to R$. The following partial differential equation describes the evolution of ϕ:

$$\frac{\partial \phi}{\partial t} = -\frac{\partial E}{\partial \phi} \qquad (6)$$

It is common to choose the signed distance function as the mapping ϕ. For a point $p \in \Omega$, this function yields the negative distance from p to the contour C if p is inside C, the positive distance from p to C if p is outside C, and 0 if $p \in C$. Replacing the contour the signed distance function phi results that $\mathbb{1}_C$ now boils down to the Heaviside function: $\mathbb{1}_C[x] = H_\phi[x] = [\phi[x] >= 0]$.

3.3 Implicit shape representation

Since ϕ is a function, optimizing over ϕ is an infinite-dimensional problem solvable with the calculus of variations. Therefore, it is beneficial to constrain ϕ to a more computationally tractable form. Cremers and Rousson (2007) proposed to implicitly represent the function ϕ using a finite set of real shape coefficients $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_M], \alpha_i \in \mathbb{R}$. Using this representation, it is now possible to explicitly model the shape coefficients $\boldsymbol{\alpha}$ in a prior $E_{shp}(\boldsymbol{\alpha}) = -\log \mathcal{P}(\boldsymbol{\alpha})$ (see Eq. 2). The authors also accounted for rigid transformations of the evolving contour (translation h, rotation θ) in their formulation. Let $G_\Phi[\boldsymbol{\alpha}] : \mathbb{R} \to \Phi$ denote a 'generator' which, for a set of shape parameters $\boldsymbol{\alpha}$ returns a signed distance function $\phi' \in \Phi$, where Φ is the domain of signed distance functions. We can then write the evolved ϕ value as (R_θ denotes the rotation matrix by angle θ):

$$\phi_{\boldsymbol{\alpha},t,\theta}(x) = G_\Phi[\boldsymbol{\alpha}](R_\theta x + t) \qquad (7)$$

The total energy $E(\boldsymbol{\alpha}, t, \theta) = E_{img}(\boldsymbol{\alpha}, t, \theta) + E_{shp}(\boldsymbol{\alpha})$ is now only a function of the shape coefficients and rigid transformation parameters and can be minimized using gradient based methods.

4. GAN-BASED SHAPE PRIORS

Generative adversarial networks (Goodfellow et al., 2014) are a class of neural networks capable of learning a mapping from a (lower-dimensional) latent variable space Z to complex, high-dimensional spaces X (e.g. images). For the purpose of generative modeling, they can be viewed as a black box $G(z) : Z \to X$ yielding an output $x \in X$ given a latent vector $z \in Z$ as input. In this work, we consider GANs as a means for generating binary images representing the shape masks of dead tree crowns (see Fig. 1a), i.e. $G(z) : [-1; 1]^{n_Z} \to 0, 1^N$. The latent variable vector z represents our shape coefficient vector $\boldsymbol{\alpha}$ (see Sec. 3.3). Usually, an input 'noise' distribution must be chosen for the latent variable space Z during the training phase of the GAN. Some popular choices include the normal and the uniform distributions. Note that the latter is particularly useful in our setting, because it would cause the shape prior probability term $\mathcal{P}(\boldsymbol{\alpha}) = \mathcal{P}(z) = \text{const}$ to become constant and hence irrelevant for the optimization of Eq. 5. Indeed, any $z \in Z$ from the valid range $[-1; 1]^{n_Z}$ will *by construction* correspond to valid shapes consistent with the training data once the GAN has been properly trained. Furthermore, the binary image output of the network means that $G(z)$ is its own indicator function, eliminating the need for an explicit $\mathbb{1}_C$ in Eq. 5. Adding the rigid transformation parameters t, θ, it suffices to optimize:

$$\begin{aligned} G_{z,t,\theta}[x] &= G(z)[R_\theta x + t] \\ E(z, t, \theta) &= E_{img}(z, t, \theta) \\ &= -\int_\Omega G_{z,t,\theta}[x] \log \mathcal{P}(\Omega_1 | I(x)) + \\ &\quad (1 - G_{z,t,\theta}[x]) \log[1 - \mathcal{P}(\Omega_1 | I(x))] dx \end{aligned} \qquad (8)$$

Since $G(z)$ is a feed-forward neural network, the gradient $\frac{\partial G(z)}{\partial z}$ can be easily obtained using the chain rule and backpropagation. Also, it should be noted that in a raster image setting, the translation values t are usually constrained to be integers (whole pixels). In order to maintain smoothness and differentiability, we use bilinear interpolation to enable $t \in \mathbb{R}$.

5. EXPERIMENT

5.1 Data acquisition

Color infrared images of the Bavarian Forest National Park, situated in South-Eastern Germany (49°3′19″ N, 13°12′9″ E),
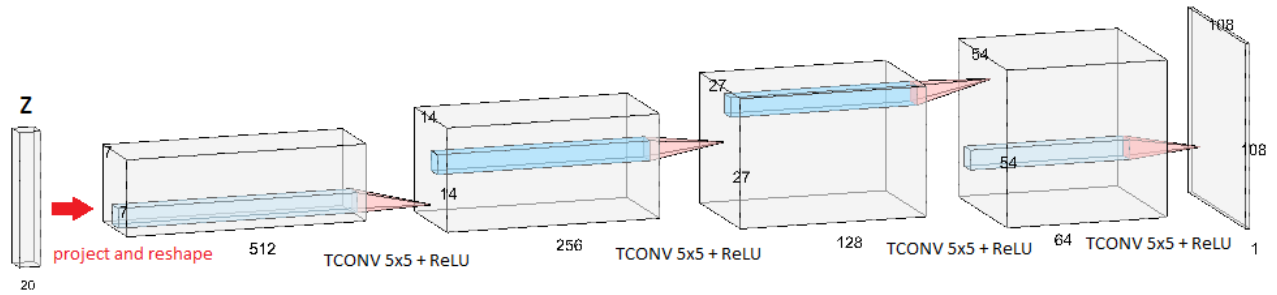
Figure 2. Architecture of the DCGAN. 'TCONV' refers to transposed convolution, whereas 'ReLU' indicates a rectified linear unit.

were acquired in the leaf-on state during a flight campaign carried out in June 2017 using a DMC III high resolution digital aerial camera.The mean above-ground flight height was ca. 2300 m, resulting in a pixel resolution of 10 cm on the ground. The images contain 3 spectral bands: near infrared, red and green.

5.2 Training and test data

We manually marked 201 outlines of dead trees within the color infrared images of a selected area in the National Park (see Fig. 3a). The distribution of their areas is depicted in Fig. 5. These manually marked were utilized for the purpose of training the semantic segmentation U-Net. We prepared patches of size 200x200 containing the input color infrared image and a pixel mask representing the labeled polygon regions. Also, we constrained the negative class labels to at most 5 pixels away from labeled dead tree polygons, to account for the fact that not all dead tree crowns in the processed images were labeled (Fig. 3b).

To train the DCGAN, we employed a different, semi automatic strategy for acquiring sample crown polygon data. We applied the trained U-Net to a new, previously unseen region of the National Park, and obtained the dead tree crown per-pixel probability map. Connected component segmentation was then applied on pixels of the image classified as dead trees. As the test area contained many overlapping and adjacent dead trees, the connected components obtained from this step usually did not represent only single trees, but rather collections of several dead tree crowns. We subsequently manually partitioned a number of connected components into individual tree crowns by applying split polylines to successively cut parts off the main polygon (Fig. 3c). We found this approach to be less time consuming than manually drawing the entire polygons. We obtained a total of 750 artificial tree crown polygons this way. They were utilized for training the DCGAN.

5.3 Evaluation criteria

We used two types of criteria to evaluate similarity between the reference and segmented polygons. First, a per-pixel similarity measure was applied: the Dice coefficient, DSC defined on two sets of pixels A, B as: $DSC(A, B) = \frac{2|A \cap B|}{|A|+|B|}$. The Dice coefficient is normalized on the interval $[0; 1]$ and measures the similarity of two sets, with a value of 0 indicating no overlap and a value of 1 indicating set equality. Second, we considered more abstract, unary properties of polygons quantifying their convexity/concavity. Specifically, we utilized the two convexity measures $c_p = p/p_c$ and $c_a = a/a_c$ described in (Jiao and Liu, 2012), where p, a, p_c, a_c denote, respectively, the target polygon's perimeter, area, as well as the perimeter and area
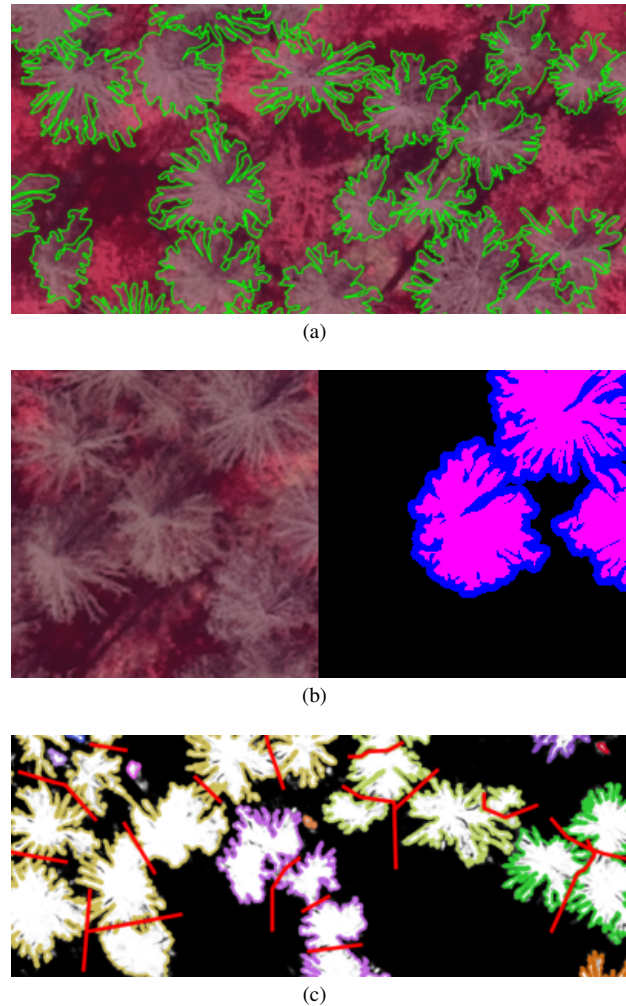

(a)

(b)

(c)

Figure 3. (a) manually marked dead tree crown polygons within CIR image, (b) left: input CIR image patch for U-Net training, right: matching label mask: positive and negative class labels indicated by magenta and blue, respectively, black mask regions do not count toward training loss, (c) manually drawn split polylines (in red) to separate connected components into individual tree crowns.

of the polygon's convex hull. We compare the measures c_p, c_a between the reference polygons and their detected counterparts to quantify the relative difference in convexity. We also calculated mean differences in the reference vs. detected polygon area and perimeter. Denoting the reference and detected polygons respectively as P_R, P_D, we define the relative differences

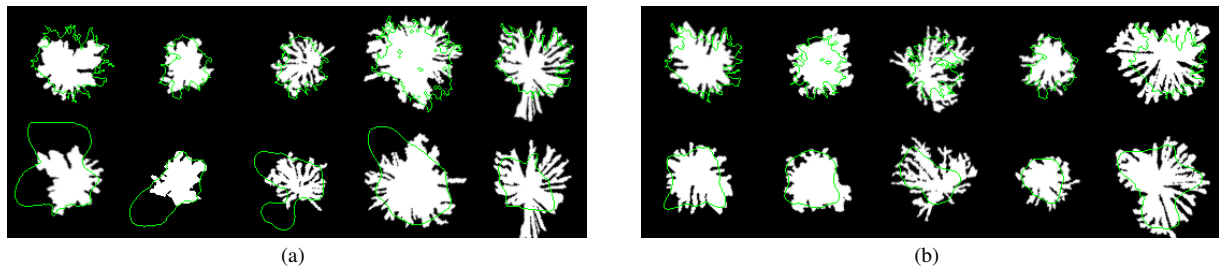(a)   (b)

Figure 4. Sample segmentation results where the GAN-based approach attained (a) better and (b) worse values of the Dice coefficient. Top and bottom rows contain, respectively, polygons reconstructed by GAN and eigenshape models plotted as green outlines over the white ground truth polygons.)

in the convexity measures as:

$$\Delta c_p(P_R, P_D) = |c_p(P_R) - c_p(P_D)|/c_p(P_R)$$
$$\Delta c_a(P_R, P_D) = |c_a(P_R) - c_a(P_D)|/c_a(P_R)$$

(9)

By analogy, we define relative differences in area and perimeter as $\Delta a(P_R, P_D), \Delta p(P_R, P_D)$.

5.4 CNN details

We utilized the DCGAN (Radford et al., 2016) in the role of the shape prior, due to its flexibility in choosing the dimensions of the output image as well as the latent vector Z. We used a Z dimension of 20 and an output image size of 108 x 108 pixels, which represents 10.8 x 10.8 meters in world coordinates. This was chosen to encompass most tree crowns of interest. We trained the GAN for 3500 iterations on the 750 original samples, with added flips and reflections as data augmentation. Sample shapes generated by the GAN at various stages of training are shown in Fig. 1. We extended the GitHub repository carpedm20/DCGAN-tensorflow (2016) to enable optimization of the objective given by Eq. 8. The overview of the used GAN architecture is shown in Fig. 2.

The tensorflow implementation of the U-Net provided by Akeret et al. (2017) was used for semantic segmentation and deriving target class posterior probabilities for all considered images. We used the original architecture proposed by Ronneberger et al. (2015), and trained the network for 2000 epochs on a total of 200 patches of size 200x200 pixels.

5.5 Experimental setup

To assess the capability of the proposed active contour approach for segmenting individual dead trees, we performed the segmentation for 200 sample images centered around the manually marked polygons (see Sec. 5.2), with additional padding of 6m around the center polygon. The objective function from Eq. 8 was optimized with respect to shape coefficients and rigid transformation parameters, using 100 restarts of gradient descent from random initializations. All computations were also performed for the baseline method, where the GAN shape prior was substituted with the original linear eigenshape signed distance function model introduced by Tsai et al. (2003). The baseline was trained on the same input data as the GAN-based approach. The segmented polygons were evaluated against the center polygon of the image only (overlap with other polygons would count as misclassified pixels). We computed the Dice coefficient as well as the relative differences $\Delta c_1, \Delta c_2, \Delta a, \Delta p$ and area.
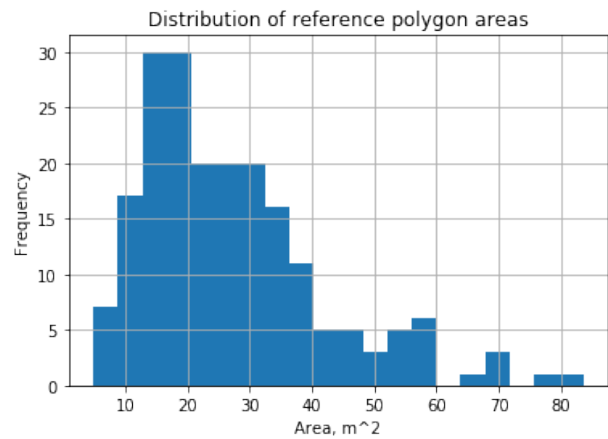

Figure 5. Distribution of areas (in square meters) associated with reference tree crown polygons.

6. RESULTS AND DISCUSSION

The numeric results of our experiments are listed in Table 1. Although the GAN and eigenshape methods achieved a nearly identical mean Coefficient of 0.69 on the whole test set, there are important differences in the behavior of the two methods. First, the test set can be partitioned into two subsets T_G, T_E such that the GAN-based method achieved better performance (measured by the Dice coefficient) on the former, whereas the eigenshape formulation was superior on the latter set (see Fig. 4 for a visual comparison of sample segmented polygons from the two sets). The performance of the GAN method is stable on both subsets (within ∓ 1 percentage points (pp) of the mean Dice coefficient), however the eigenshape variation's results deteriorate significantly on T_G, dropping by 10 pp compared to the mean. We believe this may be attributed to the average size of the reference polygons in the datasets, which is significantly smaller for T_G compared to T_E (21.9 vs 31.5 sq. m). The 20 first eigenmodes failed to capture the full variability of the training set (a total of 92% of the variance), which was dominated by polygons larger than T_G's mean value (see Fig. 5. On the other hand, it seems that the GAN-based segmentation performed better on difficult examples, since both it and the baseline degraded on T_G but the GAN's average Dice coefficient remained 9 pp higher on average.

Inspecting the images in Fig. 4, we notice that the eigenshape prior favors blob-shaped, nearly convex polygons with little fine detail. Apparently the first 20 eigenmodes of the training shapes focused on coarse details. The GAN generated images possess a more jagged boundary, with many concavities and fine details,

Data Set	Num. objects	Dice coef.	Convexity				Polygon metrics			
			Δc_p	Δc_a	$c_{p,ref}$	$c_{a,ref}$	Δp	Δa	p_{ref}	a_{ref}
All data	201				2.43	0.69			27.1	57.9
GAN		0.69	0.21	0.11			0.35	0.33		
eigenshape		0.69	0.52	0.3			0.46	0.50		
GAN superior	91				2.32	0.69			21.9	50.0
GAN		0.68	0.19	0.11			0.43	0.33		
eigenshape		0.59	0.49	0.27			0.81	0.39		
GAN inferior	110				2.51	0.69			31.5	64.0
GAN		0.70	0.22	0.10			0.29	0.33		
eigenshape		0.77	0.55	0.33			0.17	0.60		

Table 1. Comparative results between GAN and eigenshape(baseline) driven segmentation. Mean Dice coefficient, reference polygon convexity and area/perimeter ($c_{p,ref}, c_{a,ref}, a_{ref}, p_{ref}$), relative differences in convexity measures ($\Delta c_a, \Delta c_p$) and in polygon area/perimeter ($\Delta a, \Delta p$) between reference and detected polygons are shown for 3 data subsets: (1) all data, (2) only polygons where GAN attained a better Dice coefficient, (3) only polygons where GAN attained a worse Dice coefficient.

however despite moderately high Dice coefficient values, the polygons do not seem to be very well aligned with the target shapes. There are several possible reasons for this. First, examining Fig. 1, it can be seen that the GAN converged to a state where only a handful of shapes is replicated with small variations. Therefore, in some sense the GAN failed to fully learn the distribution of the training data. We hypothesize that the training set chosen through semi-automatic extraction of polygons from semantic segmentation maps might have been too homogeneous and not representative enough of all possible dead tree crown shapes. Another reason for the discrepancy between the detected and reference polygons could be convergence to weak local optima when optimizing the objective from Eq. 8. This could be addressed by utilizing a GAN that is *invertible* by design, without the need for an explicit optimization step over the latent variables, e.g. (Asim et al., 2020).

The intuitions from visual inspection of the polygons in Fig. 4 can be to some extent quantified by shape indices measuring the convexity of polygons (see Sec.5.3). On average, the GAN prior-based segmentation produces polygons which deviate by 21% in the ratio of perimeter to convex hull perimeter (c_p), and by 11% in ratio of area to convex hull area (c_a). These values are more than doubled for the eigenshape approach, at respectively 52% and 30%. Similarly, the GAN-generated polygons differ from the reference shapes by 33% and 35% in terms of perimeter and area, whereas the baseline produces polygons with an average difference of 50% and 46%. This is consistent with the more coarse and convex shapes generated by the eigenshape prior. In general, the GAN prior leads to shapes that are more similar to the target polygons with respect to all considered shape indices.

7. CONCLUSIONS

This paper presented a new formulation of active contour segmentation where the object shape prior is derived from a generative adversarial network. For an architectural choice of the GAN where the uniform distribution is used for sampling from the latent space, the optimization of the active contour objective is simplified by rendering the shape probability term constant, because all admissible latent vectors on the hypercube $[-1; 1]^{n_z}$ are valid, equal probability shapes by construction. The GAN enriched model is amenable to optimization through gradient descent since the generative component of the network

can provide gradients of the generated image with respect to the latent vector by means of backpropagation. Experiments on a real-world dataset of highly variable dead tree crown polygons showed that in some scenarios the GAN prior outperforms the eigenshape baseline in terms of per-pixel similarity of the segmented and target polygons. On average, the geometric properties of the GAN generated polygons are closer to those of the reference shapes. In this study, the GAN segmentation did not attain its full potential due to problems with convergence of the learning process and possibly a too constrained training set. An interesting future direction is to employ GANs which are invertible by construction, and which explicitly attempt to measure how well the original training data may be represented by elements of their latent space.

References

Akeret, J., Chang, C., Lucchi, A. and Refregier, A., 2017. Radio frequency interference mitigation using deep convolutional neural networks. *Astronomy and Computing* 18, pp. 35–39.

Arnab, A. and Torr, P. H. S., 2017. Pixelwise instance segmentation with a dynamically instantiated network. *CoRR*.

Asim, M., Ahmed, A. and Hand, P., 2020. Invertible generative models for inverse problems: mitigating representation error and dataset bias.

Bhattarai, N., Quackenbush, L. J., Calandra, L., Im, J. and Teale, S. A., 2012. An automated object-based approach to detect Sirex-infestation in pines. In: *ASPRS Annual Conference Proceedings*.

carpedm20/DCGAN-tensorflow, 2016. `https://github.com/carpedm20/DCGAN-tensorflow`. Accessed: 2020-05-03.

Cremers, D. and Rousson, M., 2007. Efficient kernel density estimation of shape and intensity priors for level set segmentation. In: *Deformable Models*, Topics in Biomedical Engineering. International Book Series, Springer, New York, pp. 447–460.

Cremers, D., Rousson, M. and Deriche, R., 2007. A review of statistical approaches to level set segmentation: Integrating color, texture, motion and shape. *Int. J. Comput. Vision* 72(2), pp. 195–215.

Creswell, A. and Bharath, A. A., 2019. Inverting the generator of a generative adversarial network. *IEEE Transactions on Neural Networks and Learning Systems* 30(7), pp. 1967–1974.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., 2014. Generative adversarial nets. In: Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence and K. Q. Weinberger (eds), *Advances in Neural Information Processing Systems 27*, Curran Associates, Inc., pp. 2672–2680.

Harmon, M., Franklin, J., Swanson, F., Sollins, P., Gregory, S., Lattin, J., Anderson, N., Cline, S., Aumen, N., Sedell, J., Lienkaemper, G., Jr., K. C. and Cummins, K., 1986. Ecology of coarse woody debris in temperate ecosystems. Advances in Ecological Research, Vol. 15, Academic Press, pp. 133 – 302.

He, K., Gkioxari, G., Dollar, P. and Girshick, R., 2017. Mask r-cnn. In: *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2980–2988.

Heurich, M., Ochs, T., Andresen, T. and Schneider, T., 2010. Object-orientated image analysis for the semi-automatic detection of dead trees following a spruce bark beetle (ips typographus) outbreak. *European Journal of Forest Research* 129(3), pp. 313–324.

Jensen, J. R., 2006. Remote Sensing of the Environment: An Earth Resource Perspective (2nd Edition). Prentice Hall, Upper Saddle River.

Jiao, L. and Liu, Y., 2012. Analyzing the shape characteristics of land use classes in remote sensing imagery. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* I-7, pp. 135–140.

Marcos, D., Tuia, D., Kellenberger, B., Zhang, L., Bai, M., Liao, R. and Urtasun, R., 2018. Learning deep structured active contours end-to-end. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Näsi, R., Honkavaara, E., Blomqvist, M., Lyytikäinen-Saarenmaa, P., Hakala, T., Viljanen, N., Kantola, T. and Holopainen, M., 2018. Remote sensing of bark beetle damage in urban forests at individual tree level using a novel hyperspectral camera from uav and aircraft. *Urban Forestry & Urban Greening* 30, pp. 72 – 83. Special feature: Strategic gardens and gardening: Inviting a widened perspective on the values of private green space.

Polewski, P., Yao, W., Heurich, M., Krzystek, P. and Stilla, U., 2015. Detection of single standing dead trees from aerial color infrared imagery by segmentation with shape and intensity priors. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* II-3/W4, pp. 181–188.

Polewski, P., Yao, W., Heurich, M., Krzystek, P. and Stilla, U., 2016. Combining active and semisupervised learning of remote sensing data within a renyi entropy regularization framework. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9(7), pp. 2910–2922.

Radford, A., Metz, L. and Chintala, S., 2016. Unsupervised representation learning with deep convolutional generative adversarial networks. In: Y. Bengio and Y. LeCun (eds), *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*.

Ronneberger, O., P.Fischer and Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, LNCS, Vol. 9351, Springer, pp. 234–241. (available on arXiv:1505.04597 [cs.CV]).

Safonova, A., Tabik, S., Alcaraz-Segura, D., Rubtsov, A., Maglinets, Y. and Herrera, F., 2019. Detection of fir trees (abies sibirica) damaged by the bark beetle in unmanned aerial vehicle images with deep learning. *Remote Sensing*.

Tsai, A., Yezzi, A., III, W. W., Tempany, C., Tucker, D., Fan, A., Grimson, W. and Willsky, A., 2003. A shape-based approach to the segmentation of medical imagery using level sets. *IEEE T. Med. Imaging* 22(2), pp. 137–154.

Vogelmann, J., 1990. Comparison between two vegetation indices for measuring different types of forest damage in the north-eastern united states. *Int. J. Remote Sens.* 11(12), pp. 2281–2297.

Wang, C., Lu, Z. and Haithcoat, T. L., 2007. Using Landsat images to detect oak decline in the Mark Twain National Forest, Ozark Highlands. *Forest Ecol. Manag.* 240(1-3), pp. 70 – 78.

Wu, J., Wang, Y., Xue, T., Sun, X., Freeman, B. and Tenenbaum, J., 2017. Marrnet: 3d shape reconstruction via 2.5d sketches. In: I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett (eds), *Advances in Neural Information Processing Systems 30*, Curran Associates, Inc., pp. 540–550.

Zhao, Z., Zheng, P., Xu, S. and Wu, X., 2018. Object detection with deep learning: A review. *CoRR*.