

# Diferenciação e Segmentação de Cultivo e Espécies Invasivas em Imagens Aéreas Usando Redes Neurais Convolucionais

Alexandre A. de O. Monteiro<sup>1</sup>, Paulo C. P. Jr.<sup>2</sup>, Aldo von Wangenheim<sup>2</sup>,  
Marcelo R. Stemmer<sup>1</sup>

<sup>1</sup> Departamento de Automação e Sistemas (DAS)  
Universidade Federal de Santa Catarina (UFSC) – Florianópolis, SC – Brasil

<sup>2</sup> Departamento de Informática e Estatística (INE)  
Universidade Federal de Santa Catarina (UFSC) – Florianópolis, SC – Brasil

{amonteiro, paulocpj}@incod.ufsc.br, aldo.vw@ufsc.br, marcelo.stemmer@ufsc.br

**Abstract.** *Increased use of Deep Convolutional Networks in several applications is noticeable. These Networks are being widely used for identification, classification and segmentation of objects in images, that is applied in agriculture, self-driving cars, security systems, between others. In this work, we present the result of the semantic segmentation of pixels, between soil, plant and invasive species, in a sugarcane crop, carried out by two convolutional networks: SegNet and UNet. We show the difference between them in relation to their architecture and some characteristics and, with that, what differs in the final result. The result of the segmentation can be used later in the mapping and detailing of the classes in the area of cultivation and becomes a key part to contribute to the improvement of the crop. In spite of the difficulty of creating a good dataset and realizing the ground truth of the images, it is shown that the technique generates good results and therefore has been increasingly adopted.*

**Resumo.** *É notório o aumento do uso de Redes Convolucionais Profundas em diversas aplicações. Estas Redes vêm sendo usadas em larga escala para identificação, classificação e segmentação de objetos em imagens, o que é aplicado na agricultura, em carros autônomos, em sistemas de vigilância, entre outros. Neste trabalho, é apresentado o resultado da segmentação semântica de pixels entre solo, planta e espécies invasivas, em um cultivo de cana-de-açúcar, realizado por duas redes convolucionais: SegNet e UNet. É mostrado a diferença entre elas em relação à sua arquitetura e algumas características e, com isso, o que difere no resultado final. O resultado da segmentação pode ser usado posteriormente no mapeamento e detalhamento das classes na área de cultivo e se torna peça chave para contribuir com a melhoria do plantio. Apesar da dificuldade de se criar um bom dataset e realizar o ground truth das imagens, mostra-se que a técnica gera bons resultados e por isso vem sendo cada vez mais adotada.*

## 1. Introdução

A automação da produção agrícola é um campo científico pesquisado em larga escala. Os primeiros trabalhos a respeito deste tema surgiram na década de 1980. Automatizar os processos de plantação, fertilização, proteção a plantas, colheita e contenção de ervas

daninhas vem se mostrando cada vez mais importante no contexto global, pois torna o trabalho humano menos intenso e melhora a precisão, gerando aumento na produtividade [Vidović et al. 2016]. Esse processo de automatização, em sua grande maioria, é feito a partir de imagens digitais e, muitas delas, são imagens aéreas.

Tecnologias de captação de imagem aérea têm se baseado cada vez mais em Veículos Aéreos Não Tripulados (VANT). Estes veículos têm menor custo em relação a outros equipamentos e são eficientes para coletar dados na resolução desejada. Além disso, o VANT possibilita que informações sejam coletadas mesmo em condições desfavoráveis, como por exemplo, dias nublados [Honkavaara et al. 2013]. Portanto, pode-se listar alguns benefícios que justificam a captação de imagens aéreas de campos de cultivo através dos VANT's: acesso amplo à plantação, imagens com alta resolução, rápida avaliação das condições de plantio e operação automatizada e de baixo custo [Sankaran et al. 2015].

Para se obter uma interpretação do que é captado nas imagens, utilizam-se as técnicas de visão computacional, que é um ramo da inteligência artificial [Gonzalez and Woods 2006]. O principal problema relacionado com métodos de aplicação da visão de máquina é a segmentação de imagens, ou seja, a separação de pixels que representam objetos específicos na imagem. A segmentação eficiente e automática é crucial para detecção de linhas de plantio, falhas, detecção de erva daninha, etc. [Tellaeche et al. 2008].

É possível detectar linhas de plantio e mapear falhas em campos de cultivo através de meios tradicionais de visão computacional e elementos como *Excess Green Index* (ExG), *Otsu* para binarização, *Hough Transform* (HT) e suas variações, incluindo imagens aéreas na escala de cinza. [Pérez-Ortiz et al. 2015, Chen and Ji 2010, Comba et al. 2015]. Para realizar as mesmas tarefas também há a técnica de *Object-based Image Analysis* (OBIA) [de Souza et al. 2017], Minimização Global de Energia [Vidović et al. 2016] e estratégias mais clássicas baseadas em textura [Guijarro et al. 2011] e as que trabalham em outro espaço de cores, como o HSV (Hue, Saturation, Value) e utilizam operações morfológicas em conjunto [Hamuda et al. 2017].

Sobre os métodos que se enquadram no ramo da inteligência artificial pode-se destacar a utilização de árvores de decisão através de valores NDVI (*Normalized Difference Vegetation Index*) [Delgado et al. 2012] e classificadores RNA (Rede Neural Artificial) e SVM (*Support Vector Machine*) [Andrade et al. ]. Além disso, tem crescido bastante a adoção de técnicas que se enquadram dentro do Aprendizado de Máquina, como por exemplo as Redes Neurais Convolucionais Profundas usadas para detectar plantas daninhas [Ferreira 2017].

O mapeamento de planta daninha se torna uma das tarefas mais importantes e úteis no monitoramento e gestão de fazendas [Lambert et al. 2018], e portanto, deve ser feito de forma precisa para que se possa corrigir locais específicos onde há daninhas e propiciar um cultivo eficaz. O mapeamento contribui para conter o uso indiscriminado de herbicidas, melhorando a saúde (de produtores e consumidores), reduzindo danos ao meio ambiente e ainda impede que sejam criadas daninhas cada vez mais resistentes [Louargant et al. 2018].

Pode-se perceber que há uma gama enorme de técnicas de visão computacional e diversas aplicações no ramo agrícola, seja para segmentação de objetos, detecção de falhas na plantação ou linhas plantio, que culminam em um grande auxílio para que o produtor possa aumentar sua produtividade. Contudo, percebe-se que ainda são poucos os trabalhos que realizam segmentação semântica de pixels aplicado em cultivos a partir de imagens RGB. Isso porque o uso de câmeras multi espectrais facilita bastante a separação entre solo e plantação, especialmente usando informação infra-vermelha (NIR) [Louargant et al. 2018]. Entretanto, o custo de câmeras multiespectrais é elevado em relação a câmeras tradicionais, o que muitas vezes se torna uma limitação para produtores e pesquisadores.

Portanto, a partir do que foi exposto, o objetivo deste trabalho é utilizar Redes Neurais Convolucionais para segmentar e diferenciar classes de objetos (solo, planta, espécie invasiva) em imagens RGB de uma plantação de cana-de-açúcar.

O presente trabalho é organizado da seguinte forma: na seção dois é apresentada uma revisão da literatura. Na seção três é apresentado o método proposto e seus procedimentos. Em seguida, na seção 4, são mostrados e discutidos os resultados obtidos pela pesquisa e por fim, a conclusão sobre o trabalho realizado.

## 2. Trabalhos Correlatos

A abordagem desse trabalho se baseia na união dos trabalhos relacionados com detecção e diferenciação de objetos em cultivos através de imagens aéreas e trabalhos que adotam a segmentação semântica de pixels em geral.

### 2.1. Técnicas Tradicionais de Visão Computacional

Oliveira et al. (2018) mostraram um método para detecção de falhas em uma plantação linear de café. Utilizam operadores morfológicos como *closing*, *opening* e *Hough Line Transform* (HLT) em imagens coletadas por drones. A técnica retorna a posição e o tamanho total das falhas [Oliveira et al. 2018].

Em seu trabalho, Pérez-Ortiz et al. (2015) utilizam o índice de vegetação *Excess Green*, muito usado na agricultura de precisão para propor um método de detecção de linhas de plantio. Além disso, se baseiam também no algoritmo de Otsu para binarização da imagem e *Hough Line Transform* com parâmetros otimizados para detecção das linhas, da mesma forma que Oliveira et al. (2018) e outros autores já citados. Seu trabalho contribui para simplificar a separação entre planta, solo e daninha em processos posteriores.

Com a segmentação semântica realizada no presente trabalho é possível também, em um pós-processamento, detectar as linhas de plantio, seu tamanho, e consequentemente, as falhas na plantação juntamente com suas medidas.

### 2.2. Redes Convolucionais Profundas em Problemas Gerais de Segmentação

Redes neurais convolucionais (CNNs) têm sido utilizadas em muitas tarefas diferentes de processamento de imagens e visão computacional. Quando se trata da interpretação de imagens, podemos diferenciar 3 categorias de ações: (a) classificação de imagens, (b) detecção e localização de objetos em imagens e (c) segmentação de objetos em imagens de acordo com a sua classe ou categoria. Esta última modalidade de aplicação de CNNs

é denominada *Segmentação Semântica* e é modalidade que é útil para o nosso problema. Redes para segmentação semântica classificam objetos em imagens e são capazes de associar pixels individuais das imagens à classe de objeto que representam, realizando na prática uma segmentação da imagem de acordo com a semântica do objeto ao qual cada pixel individual está associado.

Para isso essas redes neurais possuem uma arquitetura bastante típica, composta de uma etapa de classificação ou *codificação* e uma etapa de associação de rótulos a pixels ou *decodificação*, onde classificações são mapeadas de volta a pixels através de *upsampling*. Dependendo da arquitetura em particular, determinadas camadas das duas etapas da rede possuem conexões diretas ou *atalhos*. Esta estrutura é tipicamente representada em desenhos esquemáticos por um "U".

Uma das redes que foram utilizadas neste trabalho é a Unet. A sua arquitetura atinge resultados muito bons em várias aplicações de segmentação biomédica diferentes [Ronneberger et al. 2015]. Em seu artigo, Ronneberger et al. (2015) usaram bastante a prática de *data augmentation*, que foi a principal técnica que fez com que obtivessem bons resultados, já que seu conjunto de dados era pequeno.

A parte inicial da rede (*encoder*) é igual a uma arquitetura típica de uma rede convolucional de classificação de imagens. Entretanto, não possui as últimas camadas totalmente conectadas. A modificação na arquitetura feita pela Unet, consiste em ter um grande número de canais com características extraídas conectados diretamente também à etapa de *upsampling*, como pode ser visto na Figura 1. Isso permite que a rede consiga propagar mais informações sobre contexto da imagem às camadas que estão "acima", ou seja, de maior resolução. Outras características importantes sobre a Unet são mostradas na Tabela 1, ao final desta seção.

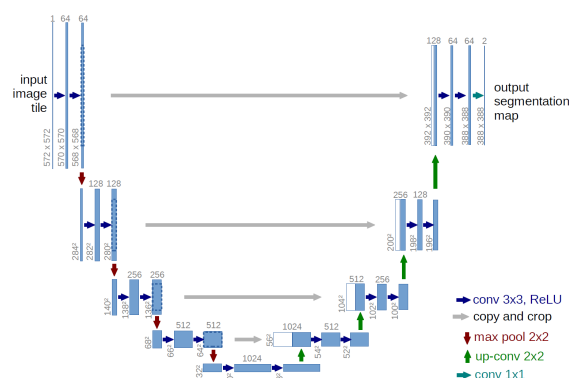


Figura 1. Topologia da rede Unet ©

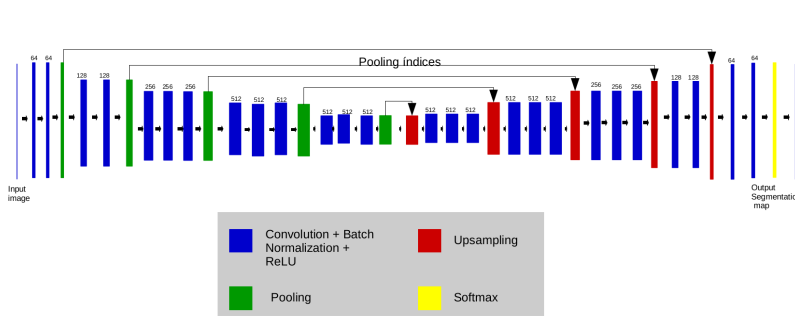
A rede SegNet, segundo Badrinarayanan et al. (2015), foi desenvolvida para corrigir um problema, que era adoção de redes convolucionais originalmente aplicadas para classificação sendo usadas para segmentação de pixels.

Na SegNet, o *encoder* é topologicamente idêntico às camadas da rede VGG13 (treze camadas de convolução) [Simonyan and Zisserman 2014], salvo as que são totalmente conectadas. Na Figura 2, é possível observar um exemplo da topologia da SegNet. Além disso, a forma como é realizado o *upsampling* é considerada o componente chave para os autores. Cada operação de *max pooling* salva um índice, que é passado posteri-

ormente às camadas apropriadas do *encoder*. As operações de *max pooling* e *upsampling* consistem basicamente em reduzir e aumentar o tamanho espacial da imagem, respectivamente. *Max pooling* é utilizado para reduzir o número de parâmetros e com isso, o esforço computacional necessário. Com o *upsampling* feito dessa forma, é possível recuperar as características mapeadas anteriormente de forma prática, rápida e ocupando menos memória, já que necessita de menos parâmetros. Isso tudo aliado à outras características mostradas na Tabela 1, tornam a SegNet eficiente e vantajosa sobre outras redes que geram resultados muito bons, porém usam mais memória e são mais lentas.

**Tabela 1. Comparativo entre as redes SegNet e UNet**

Característica	SegNet	UNet
<b>UpSampling</b>	<i>pooling</i> índices computados na etapa de <i>max pooling</i>	camadas de convolução que aprendem uma saída precisa
<b>Aplicação</b>	entendimento de cenas	processamento de imagem biomédica
<b>Camada de saída</b>	camada de classificação (softmax)	não há camadas totalmente conectadas
<b>Característica importante</b>	boa performance usando menos memória	rapidez



**Figura 2. Topologia da rede SegNet. Adaptado de Badrinarayanan et al. (2015)**

### 2.3. Segmentação Semântica em Cultivos

Segundo Sa et al. (2018), métodos inteligentes de agricultura são cada vez mais necessários para gerar aumento na produção e diminuir o impacto ambiental. Dessa forma, usando a rede neural SegNet, o objetivo foi detectar corretamente a erva daninha através da segmentação semântica de pixels, para que o tratamento não danificasse as plantas ao redor. As imagens que compõe seu *dataset* são multi espectrais captadas por um MAV (*Micro Aerial Vehicle*) e portanto, a entrada da rede conta com imagens nos canais NIR (*Near Infra Red*), *Red* e NDVI (*Normalized Difference Vegetation Index*). A vantagem é que possibilita a criação de índices de vegetação específicos baseados na taxa de radiação, que são mais robustos na variação de luminosidade. Entretanto, apesar de obter bons resultados, o conjunto de dados foi muito pequeno, o que constitui uma limitação para a análise dos resultados [Sa et al. 2018].

Uma diferença para o presente trabalho está no espectro da imagem. Enquanto Sa et al. (2018) trabalham no Infra Vermelho, neste trabalho trabalha-se com os canais RGB, pois as câmeras que operam neste espectro apresentam um custo inferior. Além disso, foram comparadas duas topologias de Rede Neural Convolutacional, SegNet e UNet, afim de analisar diferenças nos resultados obtidos. Com relação às imagens, no trabalho de Sa et al. (2018) elas foram tiradas a 2 metros de altitude. Neste trabalho, as imagens foram captadas em altitudes que podem variar entre 120 a 200 metros.

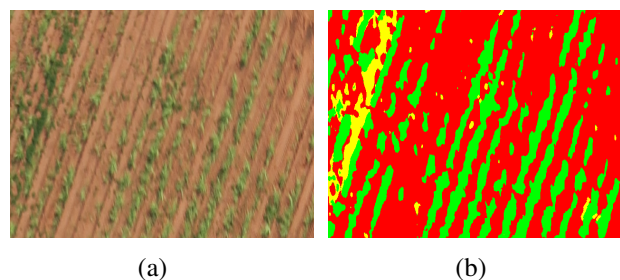
### 3. Método Proposto

O método proposto se baseia na utilização das redes neurais convolucionais profundas apresentadas na seção 2.2 para a segmentação dos pixels de imagens de uma cultura de cana-de-açúcar. Tais pixels são separados entre solo (cor vermelha), planta (cor verde) e espécie invasiva (cor amarela). Dessa forma, é possível identificar as linhas de plantio e as possíveis falhas que ocorrem na plantação.

#### 3.1. Procedimentos

Para confecção do conjunto de dados, cortou-se uma imagem de alta resolução de um talhão, resultando em 146 imagens enumeradas de 0 a 145, de 640 x 320 pixels. Desse total, foram separadas as 136 primeiras para o conjunto de treino (0 a 135) e as 10 últimas (136 a 145) para o conjunto de validação. Portanto, não houve um critério de escolha específico, apenas foram separadas priorizando-se um maior número para o conjunto de treinos, como recomendado. Imagens captadas entre 120 a 200 metros de altitude.

O processo de criação dos *labels* correspondentes às imagens foi feito manualmente, utilizando-se o software *Gimp*, por um biólogo com experiência em reconhecer diferentes espécies de plantas. O tempo para elaboração do *ground truth* foi em média de 20 minutos por imagem, totalizando cerca de 48 horas. Segue abaixo exemplo de uma imagem de entrada e seu *label* (Figura 3). Já para realizar a predição, o conjunto de teste conta com cerca de outras mil imagens de 640 x 320 pixels do mesmo talhão de cana-de-açúcar.



**Figura 3. a) Recorte de uma imagem de um talhão, b) *label* da imagem a).**

Em seguida, foram treinados os dois tipos de rede de segmentação: SegNet e UNet.

As duas redes utilizaram *Transfer Learning* com os pesos da rede VGG, treinados na base de dados ImageNet, para acelerar a fase inicial de treinamento e foram treinadas utilizando-se a API Keras, com o framework Theano como *backend*.

Segundo Chollet et al. (2015), as vantagens de se utilizar Keras são: maior produtividade e facilidade de uso, larga adoção na indústria e comunidade pesquisadora, suporte para utilização de múltiplos GPU's e adiciona praticidade na realização de testes [Chollet et al. 2015].

Os treinamentos contam com 20 épocas e 68 passos por época, *batch size* igual a 2 e taxa de aprendizagem de 0,01 com *Stochastic Gradient Descent* (SGD). A função objetivo utilizada para o treino foi a *categorical crossentropy* e como métrica de desempenho, utilizou-se *accuracy*, que mede a porcentagem de pixels corretamente classificados. O treinamento durou cerca de duas horas para cada rede rodando na GPU (*Graphics Processing Unit*). Para tal, foi utilizada uma placa de vídeo NVidia GeForce 940MX, com 2GB.

#### 4. Resultados Experimentais

Nesta seção, são mostrados os resultados de algumas imagens de saída que fazem parte do conjunto de teste. Na Figura 4, é ilustrada uma imagem para exemplificar a comparação entre o *ground truth* realizado a mão e a imagem que é gerada na camada de saída da rede.

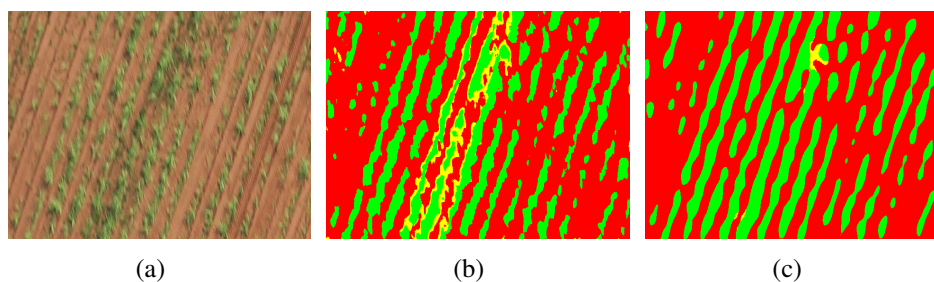


Figura 4. a) imagem de entrada ; b) rotulação ; c) predição

Na figura acima pode-se observar, na primeira coluna a imagem de entrada da rede, na segunda coluna o *label*, feito à mão e, na terceira coluna, a imagem de saída da rede.

Em seguida, na figura 5, mostra-se os gráficos obtidos durante o treinamento:

Observando os gráficos da Figura 5, pode-se notar que ambas as redes, após 20 épocas atingem uma acurácia de treinamento de 0,97, o que representa 97% do total de pixels corretamente classificados. Entretanto há uma pequena diferença, em que a UNet atinge um patamar de estabilização mais rapidamente que a SegNet. O comportamento variante da linha de acurácia do conjunto de validação é o mesmo em ambas, sendo que na UNet, o valor final é 1% maior (98%). As curvas de perda são também bem parecidas em ambas as redes.

As imagens da Figura 6, mostram a comparação dos resultados obtidos com as redes SegNet e Unet. Na coluna da esquerda um exemplo de imagem de entrada, na coluna do meio a saída da rede SegNet e na última coluna a imagem de saída da rede UNet. Estas imagens são exemplos da predição feita sobre o conjunto de teste que possui cerca de mil imagens. A partir de uma análise qualitativa, percebe-se que em relação à segmentação dos pixels que representam espécies invasivas (cor amarela), a SegNet mostra um resultado mais coeso com a imagem de entrada. Já em relação à segmentação de

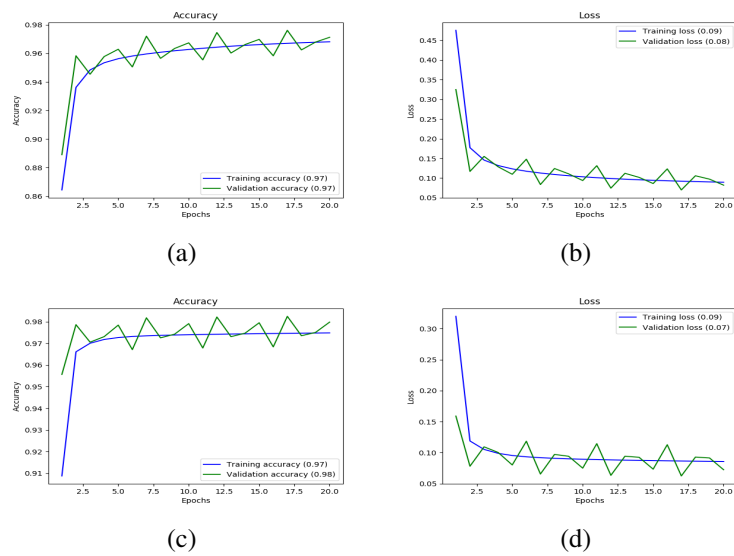


Figura 5. a) Gráfico da acurácia do modelo usando a rede SegNet nos conjuntos de treino e validação, b) Gráfico da perda nos mesmos conjuntos, c) Gráfico da acurácia do modelo usando a rede UNet nos conjuntos de treino e validação, d) Gráfico da perda nos mesmos conjuntos

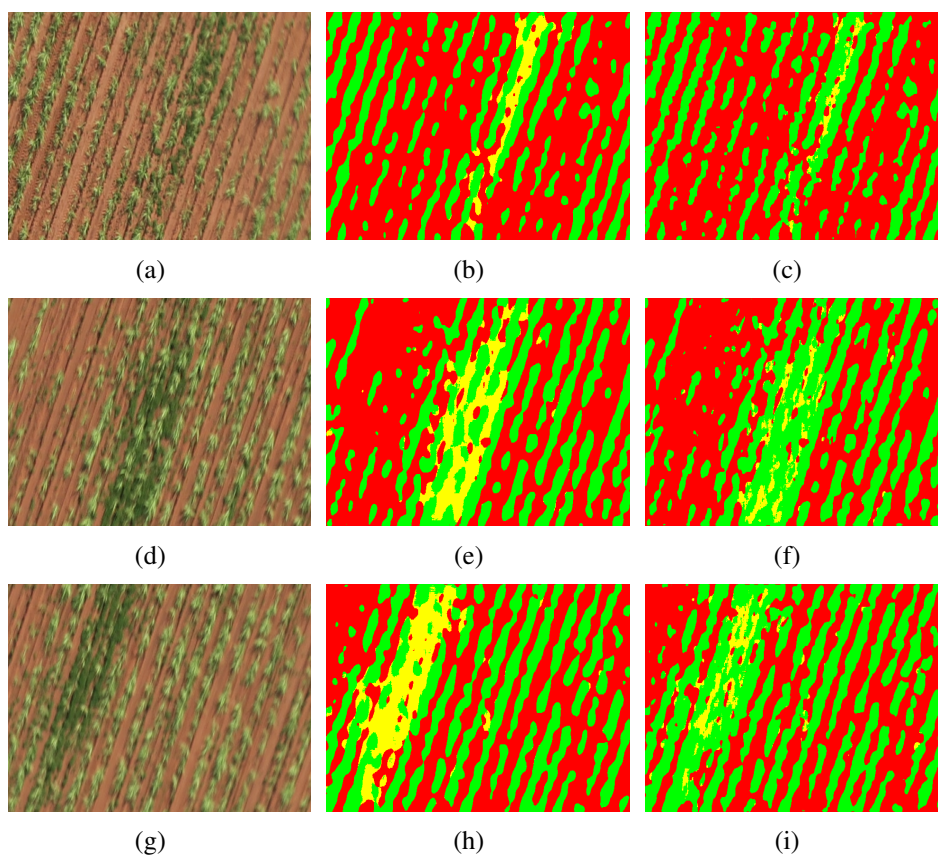


Figura 6. a) d) e) g) imagens de entrada b) e) e) h) resultados da camada de saída de rede SegNet c) f) e) i) resultados da camada de saída da rede UNet



solo e planta, ou seja, nas áreas onde não há invasivas, ambas se mostram bem assertivas e uma análise qualitativa mais minuciosa seria necessária para pontuar as diferenças.

## 5. Conclusão

Ambas as redes utilizadas neste trabalho atingiram bons resultados, para a aplicação proposta de cultivo de cana-de-açúcar, mesmo com um *dataset* de treino relativamente pequeno. O conjunto de treinamento não foi maior devido a dificuldade do processo de confecção do *ground truth*. Foi realizado um primeiro treinamento com as redes por 100 épocas e, constatou-se, que a partir da vigésima o resultado não melhora significativamente, logo, decidiu-se que vinte épocas são suficientes para gerar bons resultados.

Este trabalho mostra a segmentação semântica de pixels de uma cultura agrícola pela técnica de *Deep Learning* a partir de imagens RGB, sem a necessidade de se trabalhar com outros espectros, como por exemplo o infra-vermelho. Este fato torna-se uma vantagem já que câmeras que captam tal espectro são mais caras e menos acessíveis.

Para trabalhos futuros deseja-se realizar experimentos com um maior conjunto de dados e para tanto, maior esforço na confecção de *ground truths* será aplicado. Além disso, realizar experimentos com outras redes de segmentação promissoras e outras culturas agrícolas. Por fim, construir uma rede de segmentação própria, com características específicas a serem aplicadas para o tipo de problema em questão.

## Referências

- Andrade, L. N., Vieira, T. G., Lacerda, W. S., Davis Junior, C., Volpato, M., and Alves, H. M. Identificação automática de áreas cafeeiras em imagens de satélite utilizando redes neurais artificiais. In *Embrapa Café-Artigo em anais de congresso (ALICE)*. In.: Congresso De Pós-Graduação Da UFLA, 19, 2010, Lavras.
- Chen, L.-r. and Ji, R.-h. (2010). Detection center of the crop row by the gradient-based random hough transform [j]. *Hubei Agricultural Sciences*, 9:071.
- Chollet, F. et al. (2015). Keras. <https://keras.io>.
- Comba, L., Gay, P., Primicerio, J., and Aimonino, D. R. (2015). Vineyard detection from unmanned aerial systems images. *computers and Electronics in Agriculture*, 114:78–87.
- de Souza, C. H. W., Lamparelli, R. A. C., Rocha, J. V., and Magalhães, P. S. G. (2017). Mapping skips in sugarcane fields using object-based analysis of unmanned aerial vehicle (uav) images. *Computers and Electronics in Agriculture*, 143:49–56.
- Delgado, R. C., Sediya, G. C., Costa, M. H., Soares, V. P., and Andrade, R. G. (2012). Classificação espectral de área plantada com a cultura da cana-de-açúcar por meio da árvore de decisão. *Embrapa Monitoramento por Satélite-Artigo em periódico indexado (ALICE)*.
- Ferreira, A. d. S. (2017). Redes neurais convolucionais profundas na detecção de plantas daninhas em lavoura de soja. Master's thesis.
- Gonzalez, R. C. and Woods, R. E. (2006). *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

- Guijarro, M., Pajares, G., Riomoros, I., Herrera, P., Burgos-Artizzu, X., and Ribeiro, A. (2011). Automatic segmentation of relevant textures in agricultural images. *Computers and Electronics in Agriculture*, 75(1):75–83.
- Hamuda, E., Mc Ginley, B., Glavin, M., and Jones, E. (2017). Automatic crop detection under field conditions using the hsv colour space and morphological operations. *Computers and Electronics in Agriculture*, 133:97–107.
- Honkavaara, E., Saari, H., Kaivosoja, J., Pölönen, I., Hakala, T., Litkey, P., Mäkynen, J., and Pesonen, L. (2013). Processing and assessment of spectrometric, stereoscopic imagery collected using a lightweight uav spectral camera for precision agriculture. *Remote Sensing*, 5(10):5006–5039.
- Lambert, J., Hicks, H., Childs, D., and Freckleton, R. (2018). Evaluating the potential of unmanned aerial systems for mapping weeds at field scales: a case study with *alopecurus myosuroides*. *Weed research*, 58(1):35–45.
- Louargant, M., Jones, G., Faroux, R., Paoli, J.-N., Maillot, T., Gée, C., and Villette, S. (2018). Unsupervised classification algorithm for early weed detection in row-crops by combining spatial and spectral information. *Remote Sensing*, 10(5):761.
- Oliveira, H. C., Guizilini, V. C., Nunes, I. P., and Souza, J. R. (2018). Failure detection in row crops from uav images using morphological operators. *IEEE Geoscience and Remote Sensing Letters*.
- Pérez-Ortiz, M., Peña, J., Gutiérrez, P. A., Torres-Sánchez, J., Hervás-Martínez, C., and López-Granados, F. (2015). A semi-supervised system for weed mapping in sunflower crops using unmanned aerial vehicles and a crop row detection method. *Applied Soft Computing*, 37:533–544.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.
- Sa, I., Chen, Z., Popović, M., Khanna, R., Liebisch, F., Nieto, J., and Siegwart, R. (2018). weednet: Dense semantic weed classification using multispectral images and mav for smart farming. *IEEE Robotics and Automation Letters*, 3(1):588–595.
- Sankaran, S., Khot, L. R., Espinoza, C. Z., Jarolmasjed, S., Sathuvalli, V. R., Vandemark, G. J., Miklas, P. N., Carter, A. H., Pumphrey, M. O., Knowles, N. R., et al. (2015). Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping: A review. *European Journal of Agronomy*, 70:112–123.
- Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Tellaeché, A., Burgos-Artizzu, X. P., Pajares, G., and Ribeiro, A. (2008). A vision-based method for weeds identification through the bayesian decision theory. *Pattern Recognition*, 41(2):521–530.
- Vidović, I., Cupec, R., and Hocenski, Ž. (2016). Crop row detection by global energy minimization. *Pattern Recognition*, 55:68–86.