Regression in Convolutional Neural Networks applied to Plant Leaf Counting

Neemias Bucéli da Silva
Universidade Federal de Mato Grosso do Sul
Ponta Porã, Brasil
neemiasbsilva@gmail.com

Wesley Nunes Gonçalves
Universidade Federal de Mato Grosso do Sul
Ponta Porã, Brasil
wesley.goncalves@ufms.br

***Resumo*—Recent studies have shown that computer vision techniques developed to boost the count of plant leaves brings significant improvements. In this paper, a proposal was presented for plant leaf counting using Convolutional Neural Networks (CNNs). To accomplish the training process, CNNs architectures were adapted to solve regression problems. To evaluate the proposed method, an image dataset with 810 images of three species (*Arabidopsis*, *Tobacco* and one mutation) was used. The results showed that Xception architecture obtained the best results with R^2 of 0.96 and MAE (mean absolute error) of 0.46.**

***Index Terms*—convolutional neural networks, leaf counting, regression**

I. INTRODUÇÃO

O setor da agricultura é um dos que mais alavancam a economia no Brasil. Segundo dados estatísticos do Produto Interno Bruto (PIB) do terceiro trimestre de 2017, divulgado pelo Instituto Brasileiro de Geografia e Estatística (IBGE), o crescimento acumulado da Agropecuária no ano foi de 14,5%. Com isso, podemos concluir que a agricultura no Brasil é muito importante para a economia do país, logo se torna indispensável aprimorar manejos atuais, como também desenvolvê-los para melhorar a produtividade cada vez mais [1].

Partindo do ponto em que a agricultura de precisão é de extrema importância para estimar a produtividade, pode-se dizer que as características do genoma e fenótipo, aliado juntamente aos estudos de modelagem, possibilita a predição no desempenho das plantas sob as diversas condições ambientais [2], [3]. Para auxiliar nesta abordagem, métodos modernos de fenotipagem têm sido investidos ao longo do país, como é o caso da EMBRAPA que investiga técnicas através de imagens espectroscópicas [2]. Uma das áreas que analisa as características de imagens é a visão computacional e dentro dela existem inúmeras técnicas de fenotipagem, tal como a contagem de folhas de plantas que pode auxiliar na estimativa e eficiência da produção dos grãos [4].

Os métodos que realizam a contagem de folhas se baseiam em duas maneiras principais: obter uma segmentação por folha, que leva automaticamente ao número de folhas de uma planta [5], [6]; ou aprender o número de folhas por regressão direta a partir da imagem colorida [7]. Independente da abordagem, os métodos automáticos com visão computacional aceleram a contagem em grandes áreas quando comparado ao ser humano.

Neste artigo, é proposta uma metodologia automática para contagem de folhas em plantas observadas de cima para baixo usando CNNs. Para isso, as arquiteturas tradicionais propostas para classificação de imagens, tais como ResNet e Xception [8]–[10], foram adaptadas para regressão. Os resultados obtidos mostraram que o método proposto com a arquitetura Xception obteve um erro absoluto médio de apenas 0,46, o que torna viável o uso para contagem de folhas de plantas.

Este artigo está descrito em cinco seções. A Seção 2 apresenta os trabalhos correlatos utilizados como base para o desenvolvimento desse trabalho. Na Seção 3, é descrita a metodologia proposta para contagem automática de folhas. A Seção 4 apresenta os experimentos e resultados e, por fim, as considerações finais são apresentadas na Seção 5.

II. TRABALHOS CORRELATOS

Em 2014, a *Computer Vision Foundation* (CVF) lançou um desafio para resolver problemas de fenotipagem em forma de competição, denominada Problemas de Visão Computacional em Fenotipagem Vegetal (CVPPP). Por conseguinte, diversos algoritmos de visão computacional têm sido propostos para melhorar a acurácia dos resultados obtidos; dentro desta competição vários autores vêm propondo técnicas para contagem de folhas de plantas para melhoria na fenotipagem [7], [8], [11].

Dobrescu *et al.* [6] utilizaram CNNs e obtiveram resultados significativos como: o aumento da predição foliar utilizando o agrupamento de duas espécies de plantas diferentes (*Tobacco*, *Arabidopsis*). Eles concluíram que o treinamento em conjunto das duas espécies a partir de imagens RGB forneceu a invariância do modelo para as espécies de plantas, o que é um fator importante para a agricultura.

Shubhara *et al.* [12] propuseram uma abordagem que usa uma arquitetura denominada *SegNet*, onde essa arquitetura utiliza várias camadas de convolução seguida de camadas de *upsampling* até chegar ao tamanho da imagem original. A *SegNet* é utilizada para a segmentação inicial e depois uma outra arquitetura utilizando rede neural convolucional para a contagem de folhas. Para a contagem de folhas de plantas, foi utilizada uma estratégia de segmentação com aumento de dados.

Tabela I
RESOLUÇÃO DE CONJUNTO DE IMAGENS SEGUNDO AS RESPECTIVAS ARQUITETURAS.

ARQUITETURAS	RESOLUÇÃO
NASNet	(331x331)
ResNet50	(224x224)
InceptionResNetV2	(299x299)
Xception	(299x299)

Em um estudo recente, Jordan Ubbens *et al.* [11] apresentaram técnicas de visão computacional para a melhoria na contagem de plantas com imagens sintéticas em 3D. Os resultados mostraram que o uso de imagens sintéticas reduz o erro médio em relações aos resultados obtidos usando apenas imagens com plantas reais, que em geral estão disponíveis em um número menor.

Os trabalhos em geral usam segmentação para cada folha ou detecção de objetos. Por outro lado, esse trabalho adapta uma CNN de classificação para regressão, que não necessita de imagens segmentadas manualmente ou da posição das folhas na imagem.

III. METODOLOGIA PROPOSTA

Nesta seção é descrita a metodologia proposta para contar as folhas de plantas. Basicamente ela pode ser descrita em dois passos principais: o pré-processamento das imagens e a modelagem de uma rede neural convolucional para regressão.

A. Pré-processamento das Imagens

Como as imagens podem estar em resoluções diferentes, elas foram redimensionadas para um tamanho fixo ($W \times H$). O tamanho utilizado depende dos requisitos das arquiteturas que são utilizadas. Por exemplo, uma das arquiteturas é a Xception [13], sendo que nesta arquitetura, as imagens precisam ter resolução de 299×299. A resolução das imagens para cada arquitetura pode ser vista na Tabela I.

B. Modelagem da Rede Neural Convolucional

As CNNs são compostas por um conjunto de camadas convolucionais, ativação e *pooling* conectadas entre si. A quantidade e sequência dessas camadas são propostas na literatura como arquiteturas. As arquiteturas usadas para resolver o problema da contagem de folhas de plantas são a *Xception* [13], *ResNet* [14], *Inception* [15] e a *NasNet* [16]. Foram escolhidas de acordo com os resultados obtidos na competição *ImageNet* [17]. Para o problema deste trabalho, todas as arquiteturas utilizadas tiveram a última camada totalmente conectada modificada para um neurônio ao invés de 1000 neurônios originalmente utilizados no problema de classificação. Esse neurônio é utilizado para regressão que corresponde ao número de folhas na imagem de entrada. Um exemplo da rede neural convolucional modelada na abordagem proposta pode ser visualizada na Figura 1. Os pesos dos filtros das camadas foram inicializados de acordo com os pesos pré-treinados na *ImageNet*. As arquiteturas usadas são descritas abaixo:

- **ResNet** [14]: redes neurais mais profundas são mais difíceis de treinar. O objetivo desta arquitetura foi mostrar uma estrutura de aprendizagem residual para facilitar o treinamento das CNNs que são substancialmente mais profundas. Neste trabalho utilizamos a arquitetura com 50 camadas chamada de ResNet50 que recebe como entrada uma imagem e é composta por uma convolução com filtro (7x7) seguida de uma série de convoluções com filtros (1x1) e (3x3) terminando com uma camada totalmente conectada determinando o número da predição foliar.
- **Inception** [15]: os autores desta arquitetura buscaram diminuir a complexidade da CNN e para isso foi apresentado o módulo *Inception* que consiste em combinações paralelas de camadas com filtros convolucionais de tamanho (1x1), (3x3) e (5x5). Convoluções maiores são computacionalmente mais caras, por isso, foi proposto que sejam feitas convoluções (1x1) primeiro, para reduzir a dimensionalidade do mapa de características para depois passar por convoluções maiores.
- **Xception** [13]: a arquitetura Xception é inspirada na Inception, onde os módulos de iniciação foram substituídos por convoluções separáveis em profundidade. A Xception tem 36 camadas convolucionais formando a base de extração de características da CNN. Como a arquitetura Xception possui o mesmo número de parâmetros que a Inception [18], os ganhos de desempenho não se devem ao aumento da capacidade, mas sim ao uso mais eficiente dos parâmetros.
- **NasNet** [16]: Ao contrário das demais arquiteturas a *NasNet* é construída através de estágios recursivos chamados de blocos. Projetada para aprender qual camada ideal para o conjunto de dados de interesse. Como essa abordagem é cara quando o conjunto de dados é grande, os autores propuseram o projeto de um novo espaço de busca (conhecido como o "espaço de busca da NASNet") que possibilita a transferibilidade de aprendizado de um conjunto pequeno para um grande. Com isso, os autores obtiveram uma taxa de acurácia de 82,7% no *top-1* e 96,2% no *top-5* na *ImageNet*, sendo considerado 1,2% melhor em acurácia no *top-1*.

IV. EXPERIMENTOS E RESULTADOS

A. Base de Imagens

O banco de imagens utilizado nos experimentos pertence ao desafio de segmentação e contagem de folhas de plantas denominado LCC CVPPP2017 [7]; o conjunto de dados, conforme mostrado na Figura 2, possui 810 imagens de plantas com diferentes números de folhas. As imagens estão subdivididas em 4 pastas denominadas A1, A2, A3 e A4. Nas pastas A1 e A2, a mesma espécie foi capturada em dias diferentes. As espécies são *Arabidopsis* (pastas A1 e A2) com 159 imagens, *Tobacco* (pasta A3) com 27 imagens e uma mutação (pasta A4) com 624 imagens.

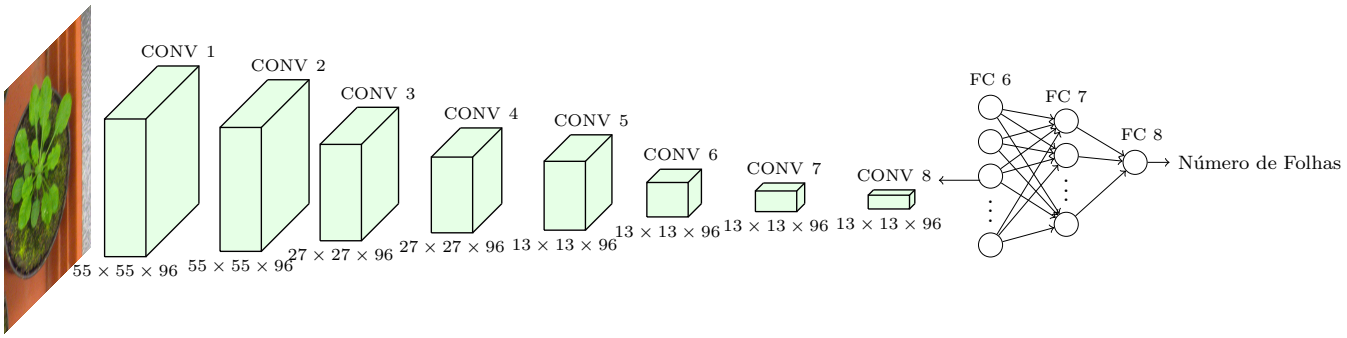


Figura 1. Rede neural convolucional para regressão utilizada na abordagem proposta.

(a) *Arabidopsis thaliana* (b) *Nicotiana tabacum* (c) *Mutação*

Figura 2. Exemplo do banco de imagens da competição CVPP2017 [7]

B. Delineamento Experimental

Todas as imagens das espécies foram divididas em 3 conjuntos de forma aleatória. O conjunto de treinamento, validação e teste possuem 60%, 20% e 20% das imagens, respectivamente. O conjunto de validação é utilizado para validar os parâmetros (número de épocas e taxa de aprendizado) utilizados nos experimentos iniciais. Finalmente, o conjunto de teste é utilizado para avaliar a predição correta obtida pela abordagem proposta. Para a execução dos treinamentos foi utilizada uma GPU. Já o otimizador escolhido no treinamento foi o RMSprop, com taxa de aprendizado (*learning rate*) atribuido em 0,0001 após experimentos empíricos no conjunto de validação. Foi definido que o treinamento seja finalizado após a CNN passar por 100 épocas. A função de perda (*loss*) utilizada foi o erro quadrático médio (*mean squared error*) onde dada imagem i e o seu número correto de folhas y_i, a função de perda L_i é determinada por:

$$L_i = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

onde n é o número de imagens e $\hat{y}_i$ é o número estimado de folhas.

C. Resultados

A Tabela II apresenta os resultados de três métricas para avaliação das arquiteturas no conjunto de treinamento e validação. As métricas abordadas são o *erro quadrático médio* (MSE - mean square error), *erro absoluto médio* (MAE - mean absolute error) e *Coeficiente de Determinação* (R^2). No conjunto de validação, a Xception obteve os melhores resultados em todas as métricas com MSE, MAE e Coeficiente de Determinação de 1,09, 0,46 e 0,96, respectivamente. Esses resultados foram seguidos pelos resultados da Inception, a segunda melhor arquitetura. Apesar de obter resultados promissores, as arquiteturas ResNet50 e NASNet forneceram resultados inferiores às duas arquiteturas anteriores. A tabela também apresenta os resultados no conjunto de treinamento para avaliar o sobreajuste (*overfitting*). Como esperado, os resultados no conjunto de validação é um pouco inferior aos resultados no conjunto de treinamento, sugerindo que houve pouco sobreajuste durante o treinamento.

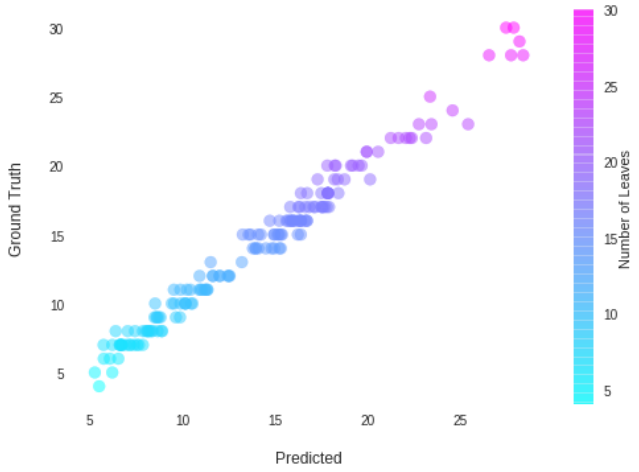
Na Figura 3 apresentamos o gráfico da predição no conjunto de teste utilizando as arquiteturas Xception, ResNet50, InceptionResNetV2 e NasNet. Nesse gráfico, o eixo x corresponde a predição enquanto que o eixo y corresponde ao número de folhas. Podemos verificar que existe uma correlação entre a predição e o número de folhas em todas as arquiteturas. Além disso, podemos observar que a abordagem proposta é capaz de predizer o número de folhas com precisão mesmo para imagens com mais de 25 folhas (canto superior direito dos gráficos).

Para entender as plantas mais difíceis que a Xception estava errando e acertando, foi realizada uma análise no conjunto de teste como mostra a Figura 4. Nas Figuras 4(a), 4(b) e 4(c) são apresentadas as três piores predições da Xception, cujo o maior erro é de aproximadamente 3 folhas. Os maiores erros acontecem quando as folhas ocupam grande parte da imagem. Já as Figuras 4(d), 4(e) e 4(f) apresentam as três melhores predições em que os erros são menores que 1.

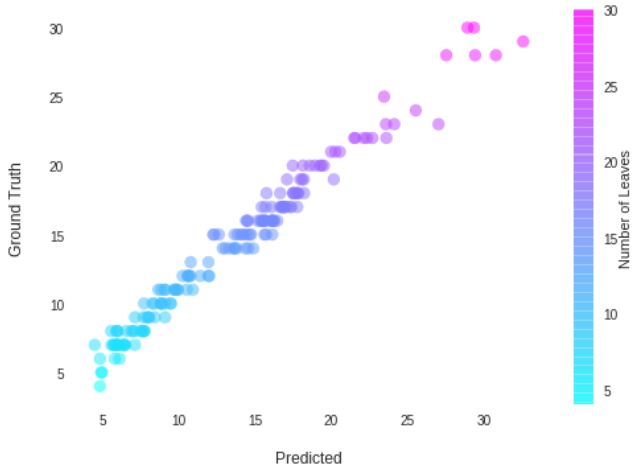
Como a Xception teve o melhor desempenho entre as arquiteturas, ela foi utilizada para contar as folhas das imagens das três espécies separadamente conforme mostra a Tabela III. Pode-se notar que onde a Xception obteve o melhor desempenho foi na Mutação, justamente, porque o número de imagens contido é relativamente satisfatório. Em contrapartida,

Tabela II
APRESENTAÇÃO DAS MÉTRICAS - ERRO QUADRÁTICO MÉDIO (MSE), ERRO ABSOLUTO MÉDIO (MAE) E COEFICIENTE DE DETERMINAÇÃO (R^2)- PARA VERIFICAR AS ARQUITETURAS ATRAVÉS DO CONJUNTO DE TREINO E VALIDAÇÃO.

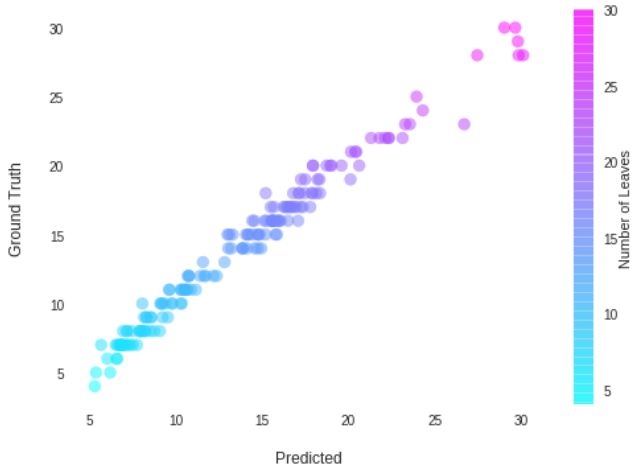
	TREINAMENTO			VALIDAÇÃO		
ARQUITETURA	**MSE**	**MAE**	$\mathbf{R^2}$	**MSE**	**MAE**	$\mathbf{R^2}$
NASNet	2,35	1,09	0,92	3,34	1,35	0,88
ResNet50	0,69	0,67	0,97	2,21	0,88	0,92
InceptionResNetV2	0,16	0,28	0,99	1,41	0.53	0.95
Xception	0,02	0,09	0,99	1,09	0,46	0,96



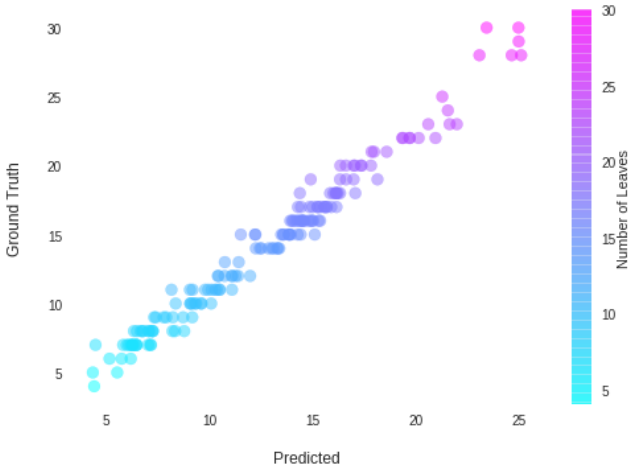
(a) Xception



(b) ResNet50

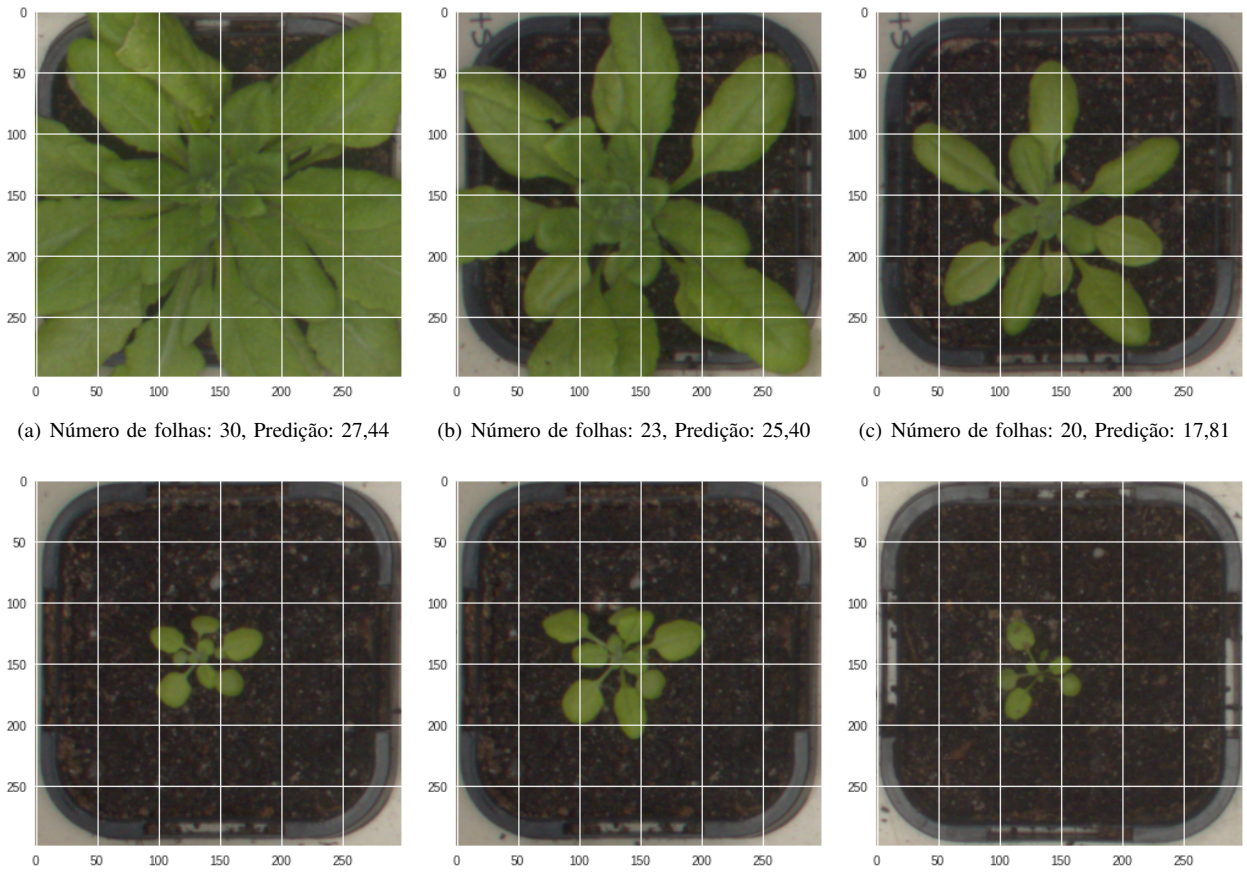


(c) InceptionResNetV2



(d) NasNetLarge

Figura 3. Predição da CNN utilizando as arquiteturas Xception, ResNet50, InceptionResNetV2 e NasNetLarge. A classe verdadeira está fixada no eixo das ordenadas denominada como "*Ground Truth*"; no eixo das abscissas está a classe que as CNNs predizeram, também está denominado como "*Predicted*".



(a) Número de folhas: 30, Predição: 27,44 (b) Número de folhas: 23, Predição: 25,40 (c) Número de folhas: 20, Predição: 17,81

(d) Número de folhas: 11, Predição: 10,98 (e) Número de folhas: 12, Predição: 12,01 (f) Número de folhas: 8, Predição: 8,03

Figura 4. Predição da *Xception* para as três piores e três melhores imagens.

a espécie *Tobacco* obteve o pior resultado por ter um conjunto de imagens pequeno. Com isso, constata-se que quanto mais imagens cada espécie tem, melhor será a avaliação dos resultados das métricas.

V. CONCLUSÃO

Neste artigo foi apresentada uma metodologia para a contagem de folhas de plantas da espécie *Arabidopsis* e o *Tobacco* utilizando uma rede neural convolucional. Os resultados mostraram que o uso da arquitetura Xception apresentou resultados significativos. Como trabalhos futuros, nós desejamos utilizar o aumento de dados para melhorar a predição foliar e comparar a abordagem proposta por Shubhara *et al.* [12]. Além disso, pretende-se usar a segmentação de folhas como pré-processamento da imagem para auxiliar na contagem posterior.

AGRADECIMENTOS

Agredecemos a FUNDECT - Fundação de Apoio ao Desenvolvimento do Ensino, Ciência e Tecnologia do Estado de Mato Grosso do Sul; ao CNPQ - Conselho Nacional de Desenvolvimento Científico e Tecnológico; a UFMS - Fundação Universidade Federal de Mato Grosso do Sul e a NVIDIA pela placa de vídeo GeForce GTX TITAN X doada e utilizada nesse trabalho.

REFERÊNCIAS

[1] Ministério da agricultura, pecuária e abastecimento. agricultura: agropecuária puxa o pib de 2017. http://www.agricultura.gov.br/noticias/agropecuaria-puxa-o-pib-de-2017. 2018.

[2] C. A. F.de SOUSA. Fenotipagem de plantas: As novas técnicas que estão surgindo para atender aos desafios atuais e futuros, 2014.

[3] Frederick B. Churchill. Wilhelm johannsen's genotype-phenotype distinction. *Journal of the History of Biology*, 7:5–30, 1974.

[4] M. Minervini, H. Scharr, and S. A. Tsaftaris. Image analysis: The new bottleneck in plant phenotyping [applications corner]. *IEEE Signal Processing Magazine*, 32(4):126–131, July 2015.

[5] Mengye Ren and Richard S. Zemel. End-to-end instance segmentation and counting with recurrent attention. *CoRR*, abs/1605.09410, 2016.

[6] Bernardino Romera-Paredes and Philip H. S. Torr. Recurrent instance segmentation. *CoRR*, abs/1511.08250, 2015.

[7] Andrei Dobrescu, Mario Valerio Giuffrida, and Sotirios A. Tsaftaris. Leveraging multiple datasets for deep leaf counting. *CoRR*, abs/1709.01472, 2017.

[8] Massimo Minervini, Andreas Fischbach, Hanno Scharr, and Sotirios A. Tsaftaris. Finely-grained annotated datasets for image-based plant phenotyping. *Pattern Recognition Letters*, 81:80 – 89, 2016.

[9] Hanno Scharr, Massimo Minervini, Andreas Fischbach, and Sotirios Tsaftaris. Annotated image datasets of rosette plants, 07 2014.

[10] Jonathan Bell and Hannah M. Dee. Aberystwyth leaf evaluation dataset, November 2016.

[11] Jordan Ubbens, Mikolaj Cieslak, Przemyslaw Prusinkiewicz, and Ian Stavness. The use of plant models in deep learning: an application to leaf counting in rosette plants. *Plant Methods*, 14(1):6, Jan 2018.

Tabela III
APRESENTAÇÃO DAS MÉTRICAS - ERRO QUADRÁTICO MÉDIO (MSE), ERRO ABSOLUTO MÉDIO (MAE) E COEFICIENTE DE DETERMINAÇÃO (R^2) - PARA VERIFICAR OS RESULTADOS DE CADA ESPÉCIE UTILIZANDO A ARQUITETURA *XCEPTION*.

ESPÉCIES	**MSE**	**MAE**	$\mathbf{R^2}$	**NÚMERO DE IMAGENS**
Tobacco	0,77	0,24	0,65	27
Arabidopsis	0,21	0,17	0,93	159
Mutação	0,24	0,11	0,99	624

[12] Shubhra Aich and Ian Stavness. Leaf counting with deep convolutional and deconvolutional networks. *CoRR*, abs/1708.07570, 2017.

[13] François Chollet. Xception: Deep learning with depthwise separable convolutions. *CoRR*, abs/1610.02357, 2016.

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

[15] Christian Szegedy, Sergey Ioffe, and Vincent Vanhoucke. Inception-v4, inception-resnet and the impact of residual connections on learning. *CoRR*, abs/1602.07261, 2016.

[16] Barret Zoph, Vijay Vasudevan, Jonathon Shlens, and Quoc V. Le. Learning transferable architectures for scalable image recognition. *CoRR*, abs/1707.07012, 2017.

[17] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.

[18] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. *CoRR*, abs/1512.00567, 2015.