Analysis of Deep Networks with Residual Blocks and Different Activation Functions: Classification of Skin Diseases

Evgin Goceri
Department of Biomedical Engineering
Engineering Faculty, Akdeniz University
Antalya, Turkey
evgin@akdeniz.edu.tr

Abstract — **Deep convolutional neural networks have been implemented for image classification tasks and achieved promising results in recent years. Particularly, ResNETs have been used since they can eliminate vanishing gradient problem in very deep networks. However, ResNET architectures with different activation functions, batch sizes, number of images in the testing and training stages can cause different results. Therefore, the effect of residual connections and activation functions image classification is still unclear. Also, in the literature, ResNET based models have been trained and tested with data sets having different characteristics. However, to make meaningful evaluations of the results obtained from different ResNET models, the same data sets should be used. Therefore, in this work, four network models have been implemented to analyze the effect of two activation functions (ReLU and SELU) and residual learning for image classification using the same data sets. To evaluate performances of these models, a real world issue, which is automated skin disease classification from colored digital images, has been handled. Experimental results and comparative analyses indicated that the ResNET with SELU and without residual block yields in the highest validation accuracy (97.01%) for image classification.**

Keywords — CNN, deep networks, image classification, ReLU, residual block, ResNET, SELU, skin diseases

I. INTRODUCTION

Image classification is a fundamental issue in computer vision and machine learning area. With advances in graphical processing unit technology, it has been observed that image classifications with deep Neural Networks (NNs) give more accurate results compared to traditional methods [1,2]. In deep NNs, input datasets are processed in deep convolutional layers, which can learn feature representations hierarchically, beginning from low level to more abstract representation. Especially, Convolutional NNs (CNNs) were proposed because they can preserve spatial relationships between image pixels [3-5].

Very deep CNNs have been usually preferred because deeper NNs have more representational power [6]. Deeper networks get this power from shallower feature representations, which are composed hierarchically, into deeper representations. However, when these networks begin to converge, a degradation problem occurs. Because, when the number of layers are increased, accuracy gets saturated and then degrades rapidly. In other words, adding more layers to a deep NN increases training error. Therefore, training of very

deep CNNs is hard due to vanishing gradients in the long forward feed and backward propagate process [7,8].

Recently, Residual NETwork (ResNET) has been applied in order to overcome vanishing gradient problems in CNNs [9,10]. A ResNET has shortcut connections parallel to the normal convolutional layers. Those shortcuts act like highways and the gradients can easily flow back. The clearest advantage of ResNETs is their fast training and convergence [9,11].

Residual blocks and chosen activation functions have a major role in the success of training of NNs. However, their effect in deep NNs is still unclear for image classification. Therefore, in this work, four network models (Table I) have been implemented to analyze the effect of residual blocks and two activation functions, which are Rectified Linear Unit (ReLU) and Scaled Exponential Linear Unit (SELU) [12].

TABLE 1. NETWORK MODELS IMPLEMENTED IN THIS WORK

Model Number	Activation Function	Batch Normalization	Residual Block
1	ReLU	Yes	Yes
2			No
3	SELU	No	Yes
4			No

Those network models designed without residual blocks (2nd and 4th model) correspond to plain networks. Therefore, in this study, performance evaluations have been performed for two plain networks and two residual connection based networks.

In this work, to evaluate accuracies of these network models in image classification, they have been applied to solve a challenging issue, which is automated classification of skin diseases from colored digital images. For this purpose, the following common skin diseases have been handled: Acne (Fig.1.a,b), Rosacea (Fig.1.c) Hemangioma (Fig.1.d), Psoriasis (Fig.1.e), Seborrheic Dermatitis (Fig.1.f).

All networks have been designed with 18 layers and 100 epochs. The learning rate is 0.00001. Also, ADAptive Moment (ADAM) estimation that is an efficient stochastic optimization has been used [13,14].

Fig. 1. Acne (a,b) [15], Rosacea (c) [16], Hemangioma (d) [17], Psoriasis (e) [18], Seborrheic Dermatitis (f) [19]

Remaining sections have been organized as follows: In section 2, a short information about ResNET architecture is given for those who are unfamiliar with ResNet. In Section 3, experimental results obtained from the network models are presented in terms of loss, validation loss, accuracy and validation accuracy. In section 4, conclusions are explained.

II. BACKGROUND: RESNET ARCHITECTURE

ResNETs consist of many residual blocks. A residual block is formulated by:

$$Y_l = h(X_l) + F(X_l, W_1) \tag{1}$$

$$X_{l+1} = f(Y_1) \tag{2}$$

In (1), the input feature of the residual block l is represented by X_l. F refers to a residual function. The term W_l, which is computed by $W_l = \{ W_{l,k} \mid_{1 \le k \le K} \}$ (K:number of layers in the residual block), corresponds to weights of the residual block l. The term f is a function, which is applied after element-wise addition. In the original ResNET [9], f is ReLU. Identity mapping is provided using the h function by $h(X_l) = X_l$. The equation given in (2) can be set into (1) to obtain the following equation, in case of the function f is an identity mapping ($X_{l+1} \equiv Y_l$);

$$X_{l+1} = X_l + F(X_l, W_l) \tag{3}$$

Therefore, the next iterative statement can be written as;

$$X_{l+2} = X_{l+1} + F(X_{l+1}, W_{l+1}) = X_l + F(X_l, W_l) + F(X_{l+1}, W_{l+1}) \tag{4}$$

Then, the following equation is obtained;

$$X_L = X_l + \sum_{i=l}^{L-1} F(X_i, W_i) \tag{4}$$

where L refers to a deeper block and l refers to a shallower block. Fig. 2 shows a residual block [10].

A. 1st Model: Network with ReLU, Batch Normalization and Residual Block

This network model has been designed with residual blocks and ReLU. Currently, it is widely-used and the most successful activation function in deep NNs due to its effectiveness and simplicity [20].

The definition of ReLU is given by f(x) = max(x, 0) [21]. If the input to the ReLU activation function is positive then gradients are able to flow. Therefore, deep networks with ReLU activation functions can be optimized more easily compared to the networks with tanh or sigmoid units.

To fix the variance and mean values of layer inputs, batch normalization, which is a non-linear transformation, has been applied for each activation.

B. 2nd Model: Network with ReLU, Batch Normalization, without Residual Block

The 2nd model has been designed without using residual blocks to see their effects. This network model has been constructed with ReLU activation function and batch normalization.
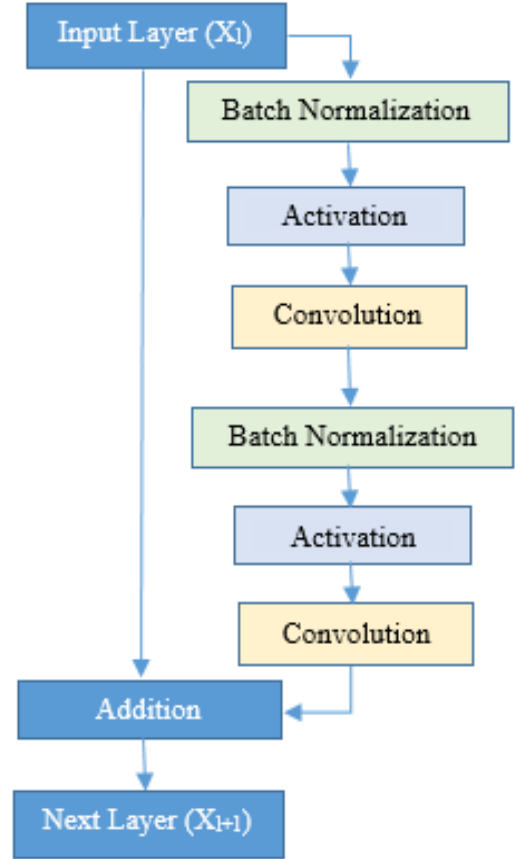

Fig. 2. Residual Block

C. 3rd Model: Network with SELU, Residual Block, without Batch Normalization

Residual blocks and SELU activation function, which was used in [22] firstly, has been applied in this model. Similar to ReLU, SELU can overcome the vanishing gradient problem in NNs. Also, in some cases, it can provide better performance compared to ReLU [23]. SELU is formulated by:

$$f(x) = \lambda \begin{cases} x & if \ x \ge 0 \\ \alpha e^{\alpha} - \alpha & if \ x < 0 \end{cases} \tag{5}$$

(with $\alpha \approx 1.7$ and $\lambda \approx 1$) and does not have batch normalization layers [20].

D. 4th Model: Network with SELU, without Batch Normalization and Residual Block

The final model has been designed using SELU activation function without batch normalization and residual block.

III. EXPERIMENTAL RESULTS

These four network models have been applied with 100 colored digital images showing skin diseases. Loss and accuracy values have been obtained from these models to evaluate their image classification performance for five common skin diseases.

In this work, balanced number of images has been used for those five classes. 70% of these total images was selected randomly for training, 15% of the total images was used for validation and the remaining 15% of the total images was used for testing.

Fig. 3 shows loss and validation loss values obtained by the 1st and 2nd model. These models have been designed with ReLU activation function. Differentiation between loss values of these two models shows the effect of the residual block used in the 1st model.
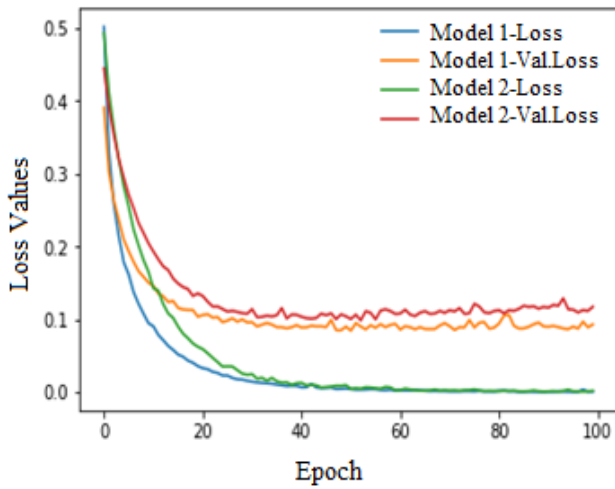
Fig.3. Loss and validation loss scores computed by the 1st model (ReLU, batch normalization and residual block) and 2nd model (ReLU and batch normalization)

Fig. 4 shows loss and validation loss values obtained by the 3rd and 4th model. These models have been designed with SELU activation function. Differentiation between loss values of these two models shows the effect of the residual block used in the 3rd model.
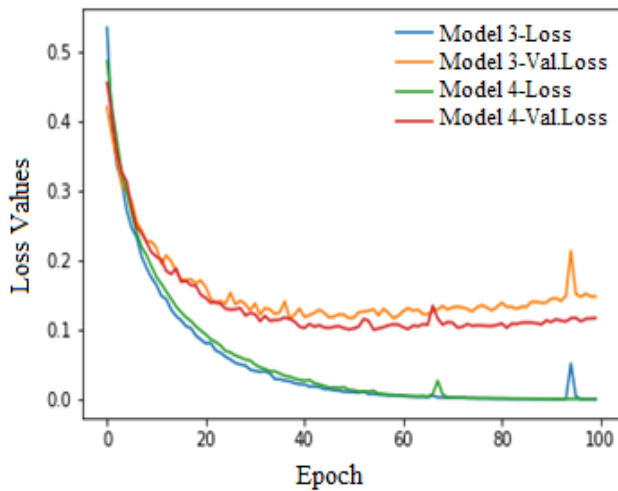

Fig.4. Loss and validation loss scores computed by the 3rd model (SELU and residual block) and 4th model (SELU)

Comparative results of the loss values obtained from the 1st and 3rd model are presented in Fig.5 to indicate the effect of ReLU and SELU activation functions.

Loss and validation loss scores computed by each model can be seen in Fig.6.

In addition to these loss values, the results indicating

accuracy of the classification have been obtained. Fig. 7 shows the accuracy and validation accuracy values obtained by the 1st and 2nd model to examine the effect of residual block used in the 1st model.
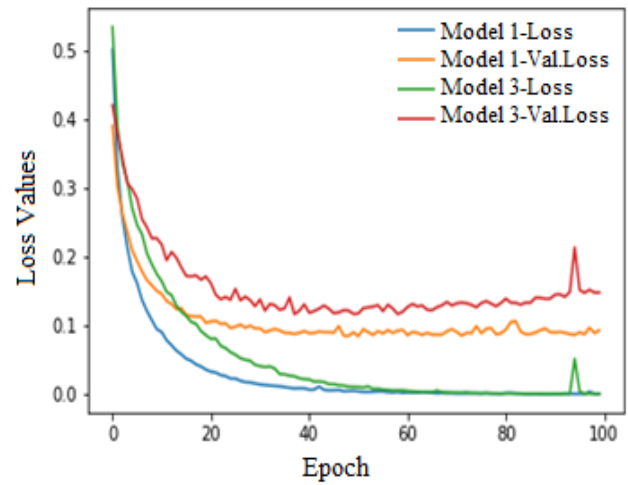

Fig.5. Loss and validation loss values scores computed the 1st model (ReLU, batch normalization and residual block) and 3rd model (SELU and residual block)
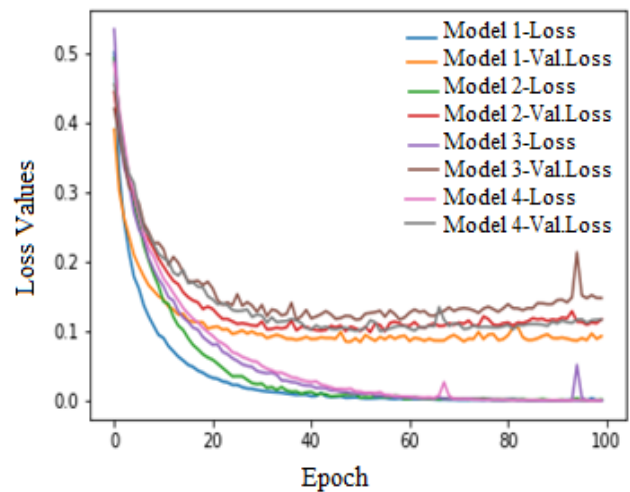

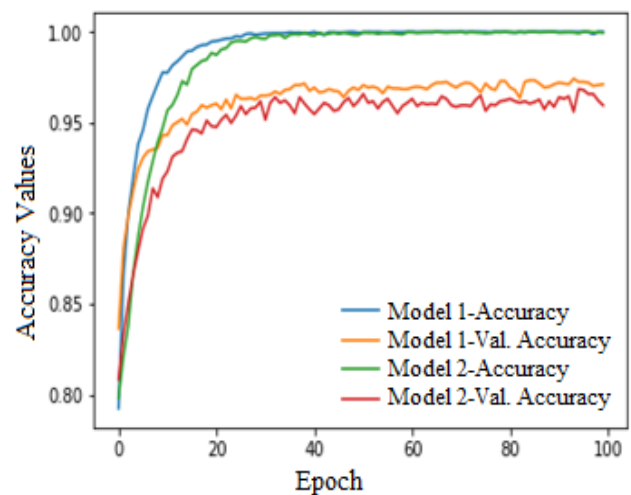Fig.6. Loss and validation loss scores computed by each model

Fig.7. Accuracy and validation accuracy values obtained by the 1st model (ReLU, batch normalization and residual block) and 2nd model (ReLU and batch normalization)

The performance in terms of accuracy of the models with residual blocks and SELU activation function (3rd and 4th model) is presented in Fig.8.
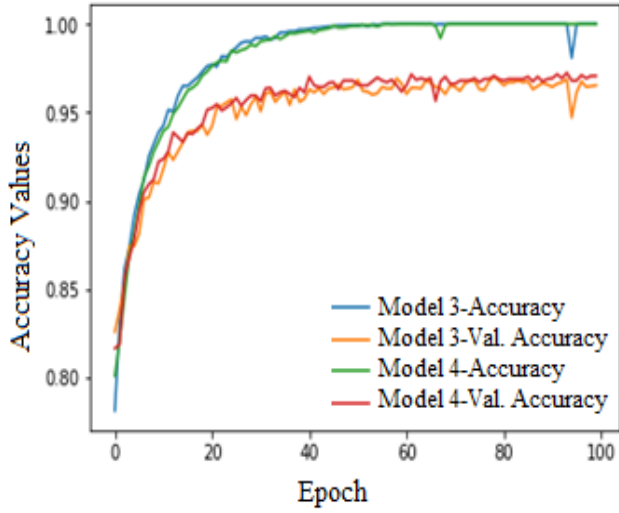

Fig.8. Accuracy and validation accuracy values obtained by the 3rd model (SELU and residual block) and 4th model (SELU)

The results obtained by the accuracy values from the 1st and 3rd model are given in Fig. 9 to indicate the effect of ReLU and SELU activation functions.
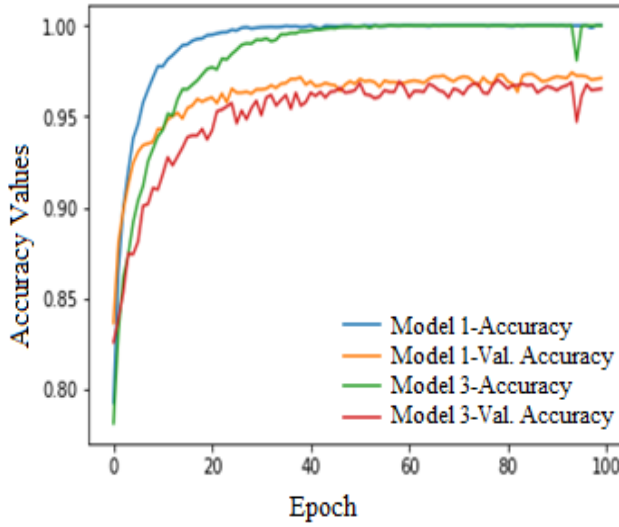

Fig.9. Accuracy and validation accuracy values obtained by the 1st model (ReLU, batch normalization, and residual block) and 3rd model (SELU and residual block)

Fig. 10 shows the accuracy and validation accuracy values obtained by each model for comparative evaluations.

The results in terms of validation loss and accuracy for each model are given as a list in Table II. It has been observed that the highest (lowest) validation accuracy is 97.01% (95.57%).
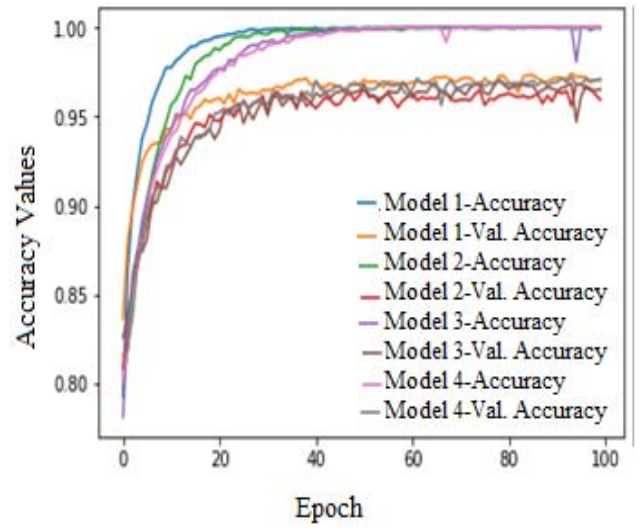

Fig.10. Accuracy and validation accuracy values of all models

TABLE II.   VALIDATION LOSS AND ACCURACY OF EACH MODELS

Models	Validation Loss	Validation Accuracy (%)
1st	0.09	96.51
2nd	0.14	95.57
3rd	0.13	96.11
4th	0.11	97.01

IV. CONCLUSION

In this work, four network models have been examined to see effects of two activation functions (ReLU and SELU) and residual blocks for image classification. Comparative analyses of these models have been performed with the results obtained by skin disease classification from colored images.

Elimination of vanishing gradient problem and self normalization with the residual blocks provides stability in training of network models (Fig.7, Fig.8).

The network models designed using SELU activation function have similar fluctuations according to the accuracy values (Fig.8).

The network model designed with SELU and without residual block gives the highest validation accuracy. On the other hand, the lowest validation loss is obtained by the network model designed with RELU and residual block (Fig.6, Fig.10).

Experimental results show that automated classification of five skin diseases can be performed with high accuracy using deep networks. These models will be tested with increased number and variation of images as an extension of this work.

REFERENCES

[1] N. D. Voulodimos, A. Doulamis and E. Protopapadakis, "Deep learning for computer vision: A brief review", Computational Intelligence and Neuroscience, vol. 2018, (Article Id: 7068349), pp.1-13, 2018.

[2] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review", Neural Computation, vol. 29, pp.2352-2449, 2017.

[3] X. Bai, B. Shi, C. Zhang, X. Cai and L. Qi, "Text/non-text image classification in the wild with convolutional neural networks", Pattern Recognition, vol. 66, pp.437–446, 2017.

[4] X. Shi, M. Sapkota, F. Xing, F. Liu, L. Cui and L. Yang, "Pairwise based deep ranking hashing for histopathology image classification and retrieval", Pattern Recognition, vol. 81, pp.14–22, 2018.

[5] Y. Zhou, Q. Hu and Y. Wang, "Deep super-class learning for long-tail distributed image classification", Pattern Recognition, vol. 80, pp.118–128, 2018.

[6] M. Telgarsky, "Benefits of depth in neural networks", The Journal of Machine Learning Research, vol. 49, pp.1–23, 2016.

[7] R. K. Srivastava, K. Greff and J. Schmidhuber, "Highway networks", In ICML 2015-Deep Learning Workshop, Lille Grande Palais, France, pp. 1-6, July 2015.

[8] K. He and J. Sun, "Convolutional neural networks at constrained time cost", In CVPR proceedings, Boston, USA, pp.5353-5360, June 2015.

[9] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition", In IEEE Conf. on Comp. Vis. Pattern Recognition proceedings, Las Vegas, USA, pp.770-778, July 2016.

[10] K. He, X. Zhang, S. Ren and J. Sun, "Identity mappings in deep residual networks", In European Conference on Computer Vision proceedings, Amsterdam, The Netherlands, pp.630-645, October 2016.

[11] C. Szegedy, S. Ioffe, V. Vanhoucke and A. Alemi. "Inception-v4. inception-resnet and the impact of residual connections on learning", In Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, USA, pp.4278-4284, February 2017.

[12] V. Nair and G.E. Hinton, "Rectified linear units improve restricted boltzmann machines", In 27th International Conference on Machine Learning proceedings, Haifa, Israel, pp.807-814, June 2010.

[13] D. P. Kingma and J. L. Ba, "ADAM: A method for stochastic optimization", In The International Conference on Learning Representations proceedings, San Diego, USA, May 2015.

[14] S. Ruder, "An overview of gradient descent optimization algorithms", arXiv:1609.04747, pp.1-14, 2016.

[15] The Website owned by The DermNet New Zelland, https://www.dermnetnz.org/topics/acne-face-images/ (Accessed on 26th May, 2019)

[16] The Website owned by The Dermnet Skin Diseases Atlas, http://www.dermnet.com/images/Rosacea (Accessed on 26th May, 2019)

[17] The Website owned by The Dermnet Skin Diseases Atlas, http://www.dermnet.com/images/Hemangioma/ (Accessed on 26th May, 2019)

[18] The Website of Skin Deep Behind The Mask, https://www.skinawareness.org/spotlight-on-psoriasis/ (Accessed on 26th May, 2019)

[19] The Website owned by The Dermnet Skin Diseases Atlas, http://www.dermnet.com/images/Seborrheic-Dermatitis (Accessed on 26th May, 2019)

[20] P. Ramachandran, B. Zoph and Q.V. Le, "Searching for activation functions", arXiv:1710.05941, pp.1-13, 2017

[21] A. Krizhevsky, I. Sutskever and G.E. Hinton, "Imagenet classification with deep convolutional neural networks", In The 26th Neural Information Processing Systems Conference proceedings, Lake Tahoe, USA, pp.1097-1105, December 2012.

[22] G. Klambauer, T. Unterthiner and A. Mayr, "Self-normalizing neural networks", In The 31st Conference on Neural Information Processing Systems proceedings, USA, pp.972-981, December 2017.

[23] T. Böhm, "A first introduction to SELUs and why you should start using them as your activation functions", Aug. 2018. (https://towardsdatascience.com/gentle-introduction-to-selus-b19943068cd9) (Accessed on 20th May, 2019)