Stabil ve stabil olmayan videoların derin evrişimli ağlarla sınıflandırılması
Classifying stable and unstable videos with deep convolutional networks

Mehmet Sarıgül[1], Levent KARACAN[2]
[1]Computer Engineering Department, Iskenderun Technical University, Hatay, Turkey
mehmet.sarigul@iste.edu.tr
[2]Computer Engineering Department, Iskenderun Technical University, Hatay, Turkey
levent.karacan@iste.edu.tr

Özet— ***Kameraların icadından bu yana, video çekmek insan için bir tutku haline gelmiştir. Ancak el kameraları, baş kameraları ve araç kameraları gibi cihazlarla kaydedilen videoların kalitesi sallantı, titreme ve istenmeyen periyodik hareketler nedeniyle düşük olabilmektedir. Video stabilizasyonu konusu onlarca yıldır çalışılmış olsa da, bir video stabilizasyon yönteminin performansının nasıl ölçüleceği konusunda bir fikir birliği yoktur. Literatürdeki birçok çalışmada, farklı yöntemlerin karşılaştırılması için farklı ölçütler kullanılmıştır. Bu çalışmada, derin evrişimli sinir ağları video stabilizasyonu için karar verici olarak kullanılmaktadır. Videoların kararlılık durumunu belirlemek için farklı sayıda katmana sahip VGG ağları kullanılmıştır. VGG ağlarının sadece iki ardışık sahne kullanarak %96,537'ye varan bir sınıflandırma performansı gösterdiği görülmüştür. Bu sonuçlar, derin öğrenme ağlarının video stabilizasyonu için bir ölçü olarak kullanılabileceğini göstermektedir.***

Anahtar Sözcükler - Video stabilizasyonu, derin öğrenme, evrişimli sinir ağları.

Abstract— **Since the invention of cameras, video shooting has become a passion for human. However, the quality of videos recorded with devices such as handheld cameras, head cameras, and vehicle cameras may be low due to shaking, jittering and unwanted periodic movements. Although the issue of video stabilization has been studied for decades, there is no consensus on how to measure the performance of a video stabilization method. In many studies in the literature, different metrics have been used for comparison of different methods. In this study, deep convolutional neural networks are used as a decision maker for video stabilization. VGG networks with different number of layers are used to determine the stability status of the videos. It was observed that VGG networks showed a classification performance up to 96.537% using only two consecutive scenes. These results show that deep learning networks can be utilized as a metric for video stabilization.**

Keywords— Video stabilization, deep learning, convolutional neural networks.

I. INTRODUCTION

Video stabilization process is the process which corrects and stabilizes the videos taken with tools such as hand cameras, head cameras, vehicle cameras, by removing shaky movements. This process can be done mechanically or optically using special hardware, or it can be performed only by using smart algorithms. All of these smart algorithms are called digital video stabilization methods [1-3]. Conventional video stabilization methods contain three steps. These steps are motion path estimation, motion path correction, and creation of stable video by processing the video images according to this new path [4,5]. With the increase in computer computing power and the development of deep learning, learning-based algorithms have begun to be recommended [6-8].

Although the issue of video stabilization has been a subject that has been studied for decades, how to determine the quality of the stable videos produced is still a matter of debate. There are several reasons for this. The first is that errors such as the classical mean square error cause blurring on the picture. Methods with very low error values can obtain very bad results visually. Except this; If the smoothing of the motion is defined as an error metric, this time the methods will correct the motion as much as possible, this time the real movements in the video may disappear. These types of situations often cause more than one error metric to be defined and used. In some studies, it was deemed appropriate for the results to be evaluated by people.

In this study, it has been shown that deep learning networks can be trained as a decision maker for video stabilization methods. The different VGG networks have been trained using only two consecutive scenes to decide whether the videos are stable or not. Trained VGG networks performed up to 96.537% in video stabilization classification. These results show that deep learning networks can be utilized as an error metric for video stabilization methods.

II. MATERIALS AND METHODS

A. Convolutional Neural Networks

Convolutional neural networks are one of the popular sub-areas of deep learning. What separates these networks from classical neural networks is the use of shared-weights in their layers. This allows both reducing the number of parameters to be trained and searching for the same pattern on the entire input image. These network structures have been widely used in many areas in the last decade [9-13].

Convolution process is carried out by predetermined fixed sized filters. These filters are passed over the entire image and activation values are calculated for each area. These calculated values are passed through an activation function and transferred to the next layer. Often, multiple convolutional layers are used consecutively to recognize more complex patterns.

B. VggNet

VggNet is one of the most popular deep learning structures [14]. The biggest reason for this popularity is the success it received in the ImageNet competition. This network structure contains many convolutional layers and is very representative. There are models such as Vgg11, Vgg13, Vgg16 ... etc., each containing a different number of layers.

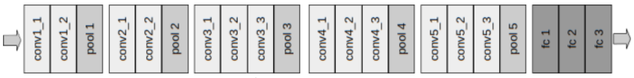

Figure 1. VggNet

C. DeepStab

DeepStab [15] is a video dataset containing 61 video pairs. These pairs consist of stable and unstable versions of the same video shot with different hardware. This dataset has been used in almost all learning based video stabilization experiments.

III. RESULTS AND DISCUSSION

Undesirable movements such as shaking and jittering are inevitable in video images taken with mobile cameras without the use of special equipment. Digital video stabilization methods aim to eliminate these unwanted camera movements and produce a stabilized video. Since video stabilization is a relative concept, it is difficult to define an error metric in this regard, so many studies have used different metrics. Moreover, in some studies, performance was measured by surveying people instead of defining a mathematical metric.

In this study, successful deep learning networks were trained to show that deep learning networks can recognize patterns that disrupt stabilization in video. Different versions of VggNet, one of the successful deep learning networks, were trained as a classifier to predict whether the videos in the DeepStab dataset are stable. 3 different VggNet networks were trained for this classification task with and without batch normalization.

Test Results	Model Classification Results	
	Model	Accuracy
	VGG11	%95.149
	VGG11 with batch normalization	%93.673
	VGG13	%96.537
	VGG13 with batch normalization	%95.948
	VGG16	%94.517
	VGG16 with batch normalization	%92.342

Table I. Classification Results

In the experiments, it was observed that deep learning networks can be trained in video stabilization as decision makers. The highest performance values, 96.537%, was achieved by Vgg13 model without batch normalization. All experiment results can be seen in Table I.

IV. CONCLUSSION

Video stabilization has been a topic that has been studied for decades. However, a common error metric could not be defined due to the multi-dimensional structure of the video stabilization problem. Some of the researchers used signal processing error metrics while others used error metrics generated for image processing problems. In some studies, it was deemed appropriate to make a decision by a user group.

In this study, it has been shown that a deep neural network can be used to define an error metric for video stabilization operation. Different VggNet networks trained with stable and unstable data showed a classification performance of up to 96.537% using only 2 consecutive pictures. This showed that deep learning networks can identify not only planar patterns but also temporal inconsistencies and can be used as an error metric for video stabilization problem.

AUTHOR CONTRIBUTIONS

All authors contributed equally to the article.

REFERENCES

[1] Matsushita, Y., Ofek, E., Ge, W., Tang, X., & Shum, H. Y. (2006). Full-frame video stabilization with motion inpainting. IEEE Transactions on pattern analysis and Machine Intelligence, 28(7), 1150-1163.

[2] Battiato, S., Gallo, G., Puglisi, G., & Scellato, S. (2007, September). SIFT features tracking for video stabilization. In *14th international conference on*

image analysis and processing (ICIAP 2007) (pp. 825-830). IEEE.

[3] Liu, S., Yuan, L., Tan, P., & Sun, J. (2014). Steadyflow: Spatially smooth optical flow for video stabilization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4209-4216).

[4] Liu, S., Tan, P., Yuan, L., Sun, J., & Zeng, B. (2016, October). Meshflow: Minimum latency online video stabilization. In *European Conference on Computer Vision* (pp. 800-815). Springer, Cham.

[5] Walha, A., Wali, A., & Alimi, A. M. (2013). Video stabilization for aerial video surveillance. AASRI Procedia, 4, 72-77.

[6] Xu, S. Z., Hu, J., Wang, M., Mu, T. J., & Hu, S. M. (2018, October). Deep video stabilization using adversarial networks. In *Computer Graphics Forum* (Vol. 37, No. 7, pp. 267-276).

[7] Wang, M., Yang, G. Y., Lin, J. K., Zhang, S. H., Shamir, A., Lu, S. P., & Hu, S. M. (2018). Deep online video stabilization with multi-grid warping transformation learning. *IEEE Transactions on Image Processing*, *28*(5), 2283-2292.

[8] Wang, M., Yang, G. Y., Lin, J. K., Shamir, A., Zhang, S. H., Lu, S. P., & Hu, S. M. (2018). Deep online video stabilization. *arXiv preprint arXiv:1802.08091*.

[9] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).

[10] Lawrence, S., Giles, C. L., Tsoi, A. C., & Back, A. D. (1997). Face recognition: A convolutional neural-network approach. *IEEE transactions on neural networks*, *8*(1), 98-113.

[11] Zhang, X., Zhou, X., Lin, M., & Sun, J. (2018). Shufflenet: An extremely efficient convolutional neural network for mobile devices. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 6848-6856).

[12] Dong, C., Loy, C. C., & Tang, X. (2016, October). Accelerating the super-resolution convolutional neural network. In *European conference on computer vision* (pp. 391-407). Springer, Cham.

[13] Xu, L., Ren, J. S., Liu, C., & Jia, J. (2014). Deep convolutional neural network for image deconvolution. In *Advances in neural information processing systems* (pp. 1790-1798).

[14] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[15] Wang, M., Yang, G. Y., Lin, J. K., Shamir, A., Zhang, S. H., Lu, S. P., & Hu, S. M. (2018). Deep online video stabilization. *arXiv preprint arXiv:1802.08091*.