Image Content Authentic Detection System using Convolutional Neural Network Method

Komang Ayu Triana Indah, Ida Bagus Putra Manuaba and Putu Manik Prihatini
Departement of Electrical Engineering, Bali State Polytechnic, Badung, Bali, Indonesia

Keywords: Fake Image, Deep Learning, Convolutional Neural Network.

Abstract: Digital content is often manipulated for a specific purpose, where image data is often falsified from the original data to provide information that is different from the original. In this study, an image detection system will be built through image classification techniques to detect image patterns that are manipulated using the Convolutional Neural Network (CNN) method. CNN is a development of Multilayer Perceptron (MLP) which is designed to process two-dimensional data. Each relationship between layers is carried out by linear operations with the existing weight values using linear convolution operations. This application serves to detect the authenticity of image content with the backpropagation process for accuracy and comparison with numbers from the training data set. The analysis of the research results produces an accuracy curve and loss validation, which states the classification of whether the image is original or has been modified. The application uses the Python programming language with Tensorflow objects to classify CNN images using two convolutional layers, one Max Pooling layer, one fully connected layer, and one output layer with softmax achieving 91.83% accuracy. Suggestions for system development, namely the use of metadata extraction with deep learning CNN can increase efficiency and reduce computational costs of the training dataset process.

1 INTRODUCTION

Social media is a means of exchanging information virtually in today's digital era. According to statistical data from the Cupo Nation portal in its statistical report, the Indonesian population is the fourth largest social media user in the world after India, the United States and Brazil. With a total of 290 million users or 19.01 percent of the total population of Indonesia. Using social media unwisely will have a negative impact on users with the development of a lot of false news in the form of content by manipulating text, image and video information using editing applications whose technology is increasingly advanced and up-to- date (Putri *et al.*, 2020).

Besides being provocative, the content has the potential to cause divisions between race, ethnicity, religion and culture in Indonesia, which is known for its diversity. Publication of digital content is often manipulated for certain purposes, where information data in the form of image data is often falsified from original data to provide information that is different from the reality. In this study, we will design a hoax detection system in the form of image content using the Convolutional Neural Network (CNN) method using the Python programming language. The input data is in the form of image content that is included in the training data set for the filtering, classification, and segmentation processes that produce output flow predictions whether the image is manipulated or not.(Li and Lyu, 2018).

2 MATERIALS AND METHOD

2.1 Electronic Information System

Electronic System is defined as a series of electronic devices and procedures that function to prepare, collect, process, analyze, store, display, announce, transmit, and/or disseminate electronic information. Therefore, there is a very close relationship between Electronic System Operators, Electronic System Users, Electronic Systems, and Electronic Information. Electronic information includes text, image, audio and video data that has been processed

Indah, K., Manuaba, I. and Prihatini, P.
Image Content Authentic Detection System using Convolutional Neural Network Method.
DOI: 10.5220/0010957400003260
In *Proceedings of the 4th International Conference on Applied Science and Technology on Engineering Science (iCAST-ES 2021)*, pages 969-975
ISBN: 978-989-758-615-6; ISSN: 2975-8246

that has meaning. Electronic data that has been processed into electronic information in its development can have positive and negative impacts that affect people's lifestyles today. The positive impact is increasing human civilization for the latest technology, while the negative impact is the spread of hoax news or lies that can affect people's mindsets.(Juditha, 2018).

2.2 Hoax Content

Hoax is misguided and dangerous information because it misleads human perception by conveying false information as truth. The impact of the spread of fake news (hoax) will have a bad impact and harm many parties, because it can cause losses from various aspects, both time and economy, public panic, worsening social relations and so on. The technique used in the manipulated information classification system is by using a machine learning-based approach. The algorithms used for text-based hoax detection include: Convolutional Neural Network (CNN), Multilayer Perceptron (MLP), Naive Bayes (NB), Random Forest (RF), Support Vector Machine (SVM) and Decision Tree (DT). (Putri *et al.*, 2019)

2.3 Digital Image Processing and Computer Vision

Image processing In the field of computers, there are actually three fields of study related to image data, but the objectives of the three are different, namely: Computer Graphics (computer graphics). Computer Graphics aims to produce images (more accurately called graphics or pictures) with geometric primitives such as lines, circles and so on. Image processing (image processing). Image Processing aims to improve image quality so that it is easily interpreted by humans or machines (in this case computers). Image processing techniques transform images into other images. So, the input is an image and the output is also an image, but the output image has a better quality than the input image. (Kinghorn, Zhang and Shao, 2018).

2.4 Convolutional Neural Network (CNN)

CNN is a type of network based on feedforward, where the information flow is only in one direction, namely from input to output. Although there are several types of CNN architectures, in general, CNNs have several convolutional layers and a pooling layer. Then, followed by one or more fully connected layers. In image classification, the input to CNN is in the form of an image, so that each pixel can be processed. In short, the convolutional layer is used as a feature extractor that learns the representation of these features from the image that is input to the CNN. Meanwhile, the pooling layer is tasked with reducing the spatial resolution of feature maps. Generally, before the fully connected layer, there is a stack of several convolutional and pooling layers that serve to extract more abstract feature representations. After that, the fully connected layer will interpret these features and perform functions that require high-level reasoning. The classification at the end of CNN will use the softmax function. (Wiriyathammabhum *et al.*, 2019).

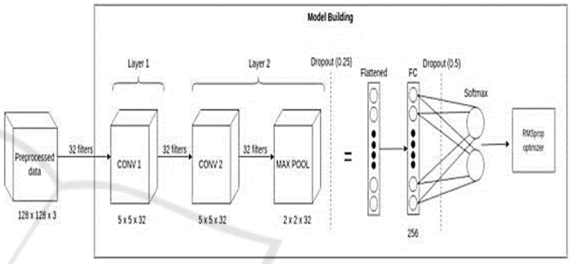

Figure 1: Convolutional Neural Network Model Architecture.

In the deep learning model used, the first layer of CNN consists of a convolutional layer with a kernel size of 5x5 and a number of filters 32. The second layer of CNN consists of a convolutional layer with a kernel size of 5x5 and a number of filters 32, and a Max Pooling layer with a size of 2x2. The two convolutional layers used use the glorot uniform kernel initializer, and the ReLU activation function to make the neurons in the convolutional layer select so that they can receive useful signals from the input data.(Hanin et al., 2019).

2.5 Deep Relationship Networks

In MTL for computer vision, the approaches often share convolutional layers, while learning the task-specific full connected layers by improving this model with Deep Relationship Networks. In addition to the shared structure and special layers, which can be seen in Figure 3, they placed the previous matrix on a fully connected layer, which allowed the model to study the relationships between tasks, similar to some Bayesian models (Kendall et al., 2018).

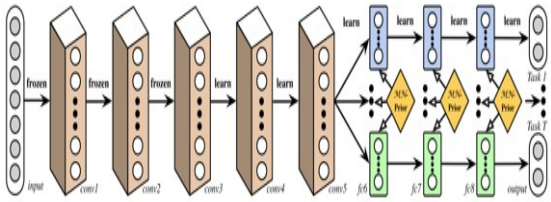

Figure 2: Deep Relationship Network with convolutional layers together and fully connected with prior matrix.

In the architecture used, only two convolutional layers are needed, because the results generated from the conversion process into an ELA image can highlight important features to determine whether an image is original or has been modified properly.

2.6 Software System Design

In designing this system, architecture and block diagrams are arranged which are divided into several processes, namely: Image Image Data Set, Preprocessing, Processed Data Set, Data Splitting, Data Loader, Model Evaluation and Load Trained Model.

System Block Diagram

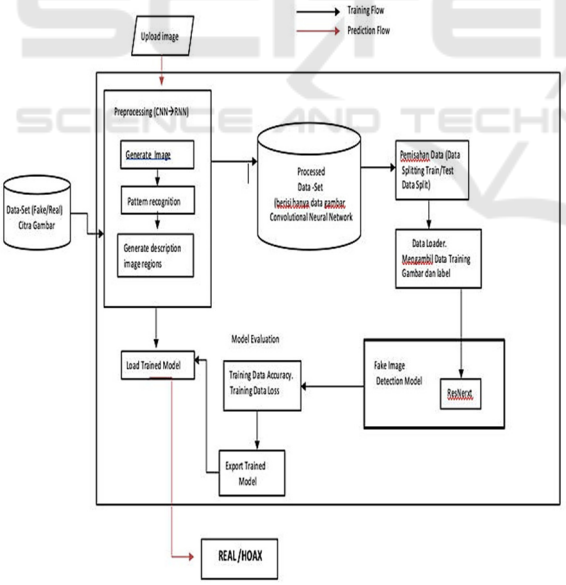

Figure 3: System Block Diagram.

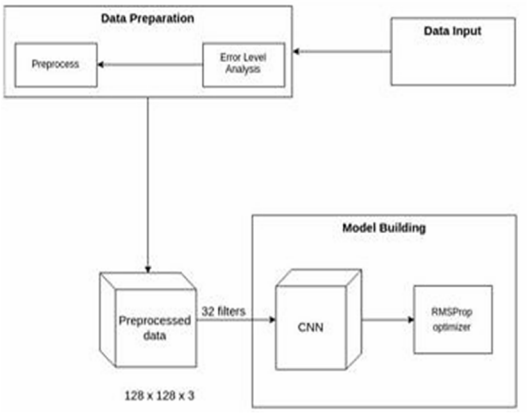

Figure 4: Convolutional Neural Network method configuration.

In general, architectural design is divided into two major parts, namely data preparation and model building. At the initial stage, input data consisting of images with the ".jpg" format, with the following details: 177 images with tampered labels and 294 images with real labels, were entered into the data preparation stage. The data preparation stage is the stage where each image which is input data is converted first into a result image. The next step is to normalize by dividing each RGB value by 255.0 to perform normalization, so that CNN converges faster (reaching the global minimum of the loss value belonging to the validation data) because the value of each RGB value only ranges between 0 and 1. The next step is to change label on a data, where 1 represents tampered and 0 represents real to be categorical value. After that, the distribution of training data and validation data was carried out using the distribution of 80% for training data and 20% for validation data. The next step is to use training data and validation data to conduct deep learning model training using CNN.

3 RESULT

3.1 Software Implementation in Software Applications

Figure 5 the following is the implementation of the application software

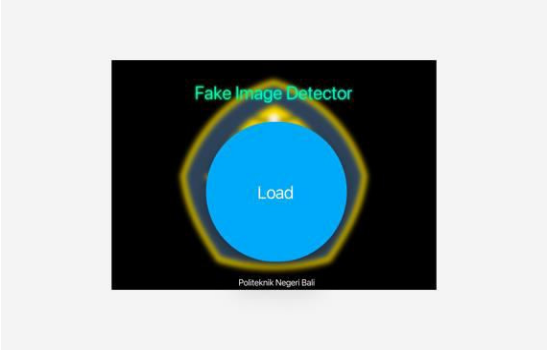

Figure 5: The main view of the manipulated image detection system architecture.

a. Input

The input is in the form of image datasets obtained from cellphone cameras/photo cameras for real data and image data from the internet/social media. The dataset is in the data preparation module which will later enter the system, for processing.

Figure 6a: Image taken directly from the phone camera (real).

Figure 6b: Edited image.

b. Process

In the next process, the image size is changed. The next step is to normalize by dividing each RGB value by 255.0 so that CNN converges faster (reaching the global minimum of the loss value belonging to the validation data) because the value of each RGB value only ranges between 0 and 1. The next step is to change the label on a data, where 1 represents tampered and 0 represents real to be categorical value. After that, the distribution of training data and validation data was carried out using the distribution of 80% for training data and 20% for validation data.

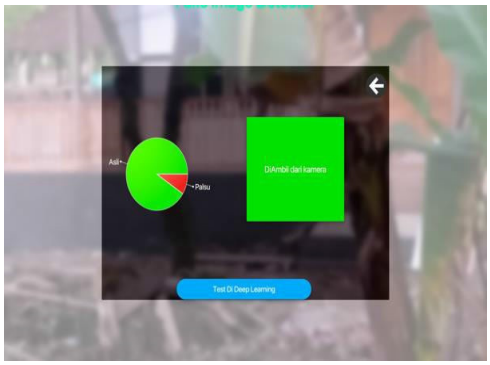

Figure 7a: Metadata Process and CNN Deep Learning (real pict).

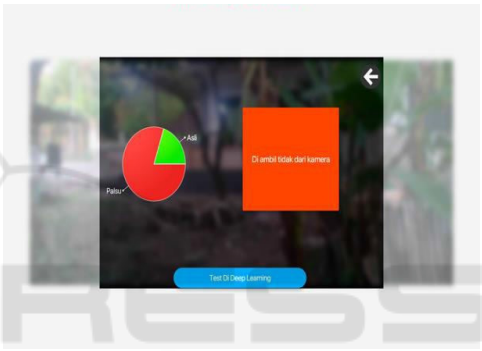

Figure 7b: CNN Metadata and Deep Learning Process (manipulated image).

Figure 8: The process of separating training data and validation data.

The next step is to use training data and validation data to conduct deep learning model training using CNN. The optimization applied during training is the RMSProp optimizer, which is one of the adaptive learning rate methods. The complete architecture used in the model building section can be seen in the image below or by using the link which is a complete architectural drawing.

c. Output

In the deep learning model used, the first layer of CNN consists of a convolutional layer with a kernel size of 5x5 and a number of filters 32. The second layer of CNN consists of a convolutional layer with a kernel size of 5x5 and a number of filters 32, and a Max Pooling layer with a size of 2x2. The two convolutional layers used use the glorot uniform kernel initializer, and the ReLU activation function to make the neurons in the convolutional layer select so that they can receive useful signals from the input data.

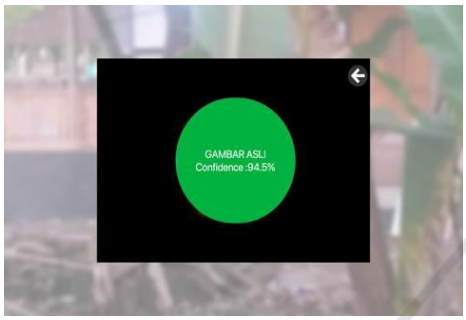

Figure 9a: Meta data extraction results for original image results.

Figure 9b: Meta data extraction results for manipulated image results.

3.2 Software Training Data Output from Data Preparation Modul

The next step is to use training data and validation data to conduct deep learning model training using CNN. The optimization applied during training is the RMSProp optimizer, which is one of the adaptive learning rate methods. The complete architecture used in the model building section can be seen in the image below or by using the link which is a complete architectural drawing.

The process of detecting fake images like the following program listing. The output of training data from metadata preparation is in the following coding listing:

```
with a Fakeness from metdata =
0.09090909
Number of metadata fields = 64
Fakeness from metdata = 0.8
Number of metadata fields = 20
```

The results of the Deep Learning test for real image metadata results are as follows:

```
Loading NN........
Loading Image
:/Users/ayutriana/Desktop/image/origi
na l.jpg
Dimension is set to
java.awt.Dimension[width=100,height=1
00
]
{faked=0.0, real=1.0}
Neural net results:-
{faked=0.0, real=1.0}
Metadata Result: Fakeness =
0.09090909 Fakeness from metdata =
0.09090909 Number of metadata fields =
64
```

The results of the Deep Learning test for fake image metadata results are as follows:

```
 Loading NN........
Loading Image
:/Users/ayutriana/Desktop/RESEARCH
DIPA 2021/image/fake_wa.jpeg
Dimension is set to
java.awt.Dimension[width=100,height=100
]
Nueral network loaded =
/Users/ayutriana/Desktop/fake_image_d
et ector/nnet/MLPV2.0.nnet
Learning Rule =
org.neuroph.nnet.learning.MomentumBackp
ropagation@1cecd71a
{faked=1.0, real=1.0}
Neural net results:-
{faked=1.0, real=1.0}
Result Metadata: Fakeness = 0.8
```

The following is the meta data information used in the training datasets:

JPEG----------

[JPEG] Compression Type - Baseline
[JPEG] Data Precision - 8 bits [JPEG] Image Height - 3024 pixels [JPEG] Image Width - 4032 pixels
[JPEG] Number of Components - 3
[JPEG] Component 1 - Y component: Quantization table 0, Sampling

factors 2 horiz/2 vert [JPEG] Component 2 - Cb component: Quantization table 1, Sampling factors 1 horiz/1 vert [JPEG] Component 3 - Cr component: Quantization table 1, Sampling factors 1 horiz/1 vert
---------- Exif IFD0----------
[Exif IFD0] Orientation - Right side, top (Rotate 90 CW)
[Exif IFD0] X Resolution - 72 dots per inch [Exif IFD0] Y Resolution - 72 dots per inch [Exif IFD0] Resolution Unit - Inch
[Exif IFD0] YCbCr Positioning - Center of pixel array
---------- Exif SubIFD----------
[Exif SubIFD] Exif Version - 2.21
[Exif SubIFD] Components Configuration - YCbCr
[Exif SubIFD] FlashPix Version - 1.00 [Exif SubIFD] Color Space - sRGB
[Exif SubIFD] Exif Image Width - 4032 pixels [Exif SubIFD] Exif Image Height - 3024 pixels [Exif SubIFD] Scene Capture Type - Standard
---------- Exif Thumbnail----------
[Exif Thumbnail] Compression - JPEG (old- style)
[Exif Thumbnail] X Resolution - 72 dots per inch
[Exif Thumbnail] Y Resolution - 72 dots per inch
[Exif Thumbnail] Resolution Unit - Inch
[Exif Thumbnail] Thumbnail Offset - 286 bytes
[Exif Thumbnail] Thumbnail Length - 5966 bytes
---------- ICC Profile----------
[ICC Profile] Profile Size - 548 [ICC Profile] CMM Type - appl [ICC Profile] Version - 4.0.0
[ICC Profile] Class - Display Device [ICC Profile] Color space - RGB
[ICC Profile] Profile Connection Space - XYZ [ICC Profile] Profile Date/Time - 2017:07:07 13:22:32
[ICC Profile] Signature - acsp
[ICC Profile] Primary Platform - Apple Computer, Inc.
[ICC Profile] Device manufacturer - APPL [ICC Profile] XYZ values - 0.964 1 0.825 [ICC Profile] Tag Count - 10
[ICC Profile] Profile Description - Display P3 [ICC Profile] Copyright - Copyright Apple Inc.,2017
[ICC Profile] Media White Point - (0.9505, 1, 1.0891)
[ICC Profile] Red Colorant - (0.5151, 0.2412, 65536)
[ICC Profile] Green Colorant - (0.292, 0.6922,0.0419)
[ICC Profile] Blue Colorant - (0.1571, 0.0666, 0.7841)
[ICC Profile] Red TRC - para (0x70617261): 32 bytes
[ICC Profile] Chromatic Adaptation - sf32 (0x73663332): 44 bytes
[ICC Profile] Blue TRC - para (0x70617261): 32 bytes
[ICC Profile] Green TRC - para (0x70617261): 32 bytes

File----------
[File] File Name - REAL camera 2.jpg [File] File Size - 1550651 bytes
[File] File Modified Date - Mon Jul 26 13:03:19 +08:00 2021

4 DISCUSSION

Based on the results of trials in this study, it can be concluded as follows: obtained from the proposed method has a maximum accuracy of 91.83%. The image of the accuracy curve and loss curve can be seen appearing 5x5 and filter af 32. The second layer of CNN consists of a convolution layer with a kernel size of 5x5 and a filter count of 32, and a Max Pooling layer with a size of 2x2. Two uses used It can be seen in the figure above that the best accuracy is obtained at epoch 0.220. The validation loss value at a linear value of 5 starts to level off and eventually increases, which is a sign of overfitting. A good method for identifying the number of epochs to use during training is early termination. With this method, the training will be stopped when the validation accuracy value starts to decrease or the validation loss value starts to increase. The test uses an image dataset obtained from data preparation which will later be processed by the system .

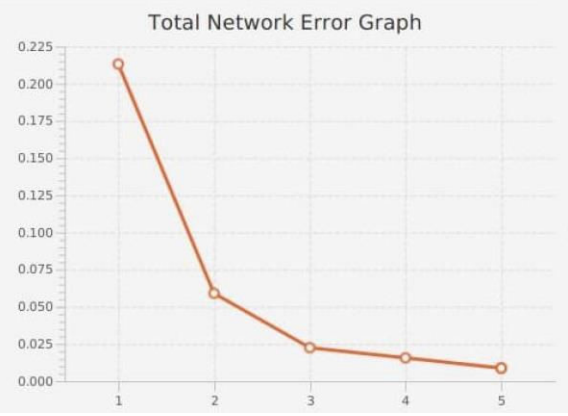

Figure 10: Accuracy curves and loss curves for training data and validation data.

The number of training epochs required is small to achieve convergence, because the use of the resulting image features makes model training much more efficient, and the normalization performed on the RGB values for each pixel also speeds up the convergence of the CNN model.

5 CONCLUSION

Based on the results of trials in this study, the following conclusions can be drawn: CNN using two convolutional layers, one MaxPooling layer, one fully connected layer, and one output layer with softmax can achieve 91.83% accuracy. The use of metadata extraction with deep learning CNN can increase efficiency and reduce the computational costs of the training process. This can be seen from the reduction in the number of layers from the previous method and the number of epochs required.

REFERENCES

Grant, J. T. (1984) 'Background subtraction techniques in surface analysis', Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films, 2(2), pp. 1135–1140. doi:10.1116/1.572689.

Lahagu, J. (2019) 'Detecting the Originality of a Digital Image By Applying the Method

Adler-32', KOMIK (National Conference on Information and Computer Technology), 3(1), pp. 789–797. doi:10.30865/komik.v3i1.1694.

Van Den Oord, A., Kalchbrenner, N. and Kavukcuoglu, K. (2016) 'Pixel Recurrent neural networks', 33rd International Conference on Machine Learning, ICML 2016, 4, pp. 2611–2620.

Research, L. and Internal, H. (2019) 'Internal Grants for News Detection Systems Fake News on Social Media'.

Zgöbek, zlem, JA Gulla, (2017). "Towards an Understanding of Fake News", Norwegian Big Data Symposium Kshetri, Nir, J. Voas, "The Economics of 'Fake News'", IT Pro (November/December 2017), IEEE Computer Society.

Chinese Academy of Sciences. "CASIA Image Tempering Detection Evaluation Database (CAISA TIDE) V2.0. Taken from http://forensics.idealtest.org

N. Krawetz (2007), "A pictures worth digital image analysis and forensics," Black Hat Briefings, p. 1-31,

Rawat, Waseem, Z. Wang, (2017), "Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review", Neural Computation 29 p. 2352-2449.

Gunawan, Teddy Surya, Hanafiah, SAM, Kartiwi, M., Ismail, N., Za'bah, NF, Nordin, AN, (2017) "Development of Photo Forensics Algorithm by Detecting Photoshop Manipulation Using Error Level Analysis", Indonesian Journal of Electrical Engineering and Computer Science, Vol. 7, No. 1, , p. 131- 137.

Photo Forensics: Detect Photoshop Manipulation with Error Level Analysis, September 2018. Taken from https://resources.infosecinstitute.com/error- level-analysis-detect-image- manipulation/#gref

Edelman, (2018) Edelman Trust Barometer Global Report", taken from https://cms.edelman.com/ sites/ default/files/201 8- 01/2018%20Edelman %20Trust %20Barometer %20Global%20Report.pdf

Bulls, Jeff (2009),"6 Powerful Reasons Why you Should include Images in your Marketing", taken from https://www.jeffbullas.com/6- powerful-reasons-why-you-should-in include- images-in-your-marketing-infographic/

Villan, M. Afsal, Kuruvilla, K., Paul, J., Elias, E. P., (2017)"Fake Image Detection Using Machine Learning", *International Journal of Computer Science and Information Technology & Security (IJCSITS)*, Vol. 7, No. 2.

V. Nair and G. E. Hinton, (2010) "Rectified linear units improve restricted boltzmann machines," Proceedings of the 27th International Conference on Machine Learning, June 21-24, p. 807-814.

Putri, T. T. A., Mendoza, M. D., & Alie, M. F. (2020). Sentiment Analysis On Twitter Using The Target-Dependent Approach And The Support Vector Machine (SVM) Method: Sentiment Analysis On Twitter Using The Target-Dependent Approach And The Support Vector Machine (SVM) Method. *Jurnal Mantik*, *4*(1), 20-26.

Juditha, C. (2018). Hoax Communication Interactivity in Social Media and Anticipation (Interaksi Komunikasi Hoax di Media Sosial serta Antisipasinya). Pekommas, 3(1), 261723.

Kinghorn, P., Zhang, L., & Shao, L. (2018). A region-based image caption generator with refined descriptions. Neurocomputing, 272, 416-424.

Wiriyathammabhum, Peratham, Abhinav Shrivastava, Vlad I. Morariu, and Larry S. Davis. "Referring to objects in videos using spatio-temporal identifying descriptions." arXiv preprint arXiv:1904.03885 (2019).

Hanin, B., & Rolnick, D. (2019). Deep relu networks have surprisingly few activation patterns. Advances in neural information processing systems, 32.

Kendall, A., Gal, Y., & Cipolla, R. (2018). Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In Proceedings of the IEEE conference on computer vision and pattern recognitio n (pp. 7482-7491).