Sentiment Analysis From Images - Comparative Study of SAI-G and SAI-C Models' Performances Using AutoML Vision Service from Google Cloud and Clarifai Platform

Daniela Marcu[†] **and Mirela Danubianu**[2††],

[1†] , [2††] Stefan cel MareUniversity of Suceava, Romania

Summary
In our study we performed a sentiments analysis from the images. For this purpose, we used 153 images that contain: people, animals, buildings, landscapes, cakes and objects that we divided into two categories: images that suggesting a positive or a negative emotion. In order to classify the images using the two categories, we created two models. The SAI-G model was created with Google's AutoML Vision service. The SAI-C model was created on the Clarifai platform.
The data were labeled in a preprocessing stage, and for the SAI-C model we created the concepts POSITIVE (POZITIV) AND NEGATIVE (NEGATIV).
In order to evaluate the performances of the two models, we used a series of evaluation metrics such as: Precision, Recall, ROC (Receiver Operating Characteristic) curve, Precision-Recall curve, Confusion Matrix, Accuracy Score and Average precision.
Precision and Recall for the SAI-G model is 0.875, at a confidence threshold of 0.5, while for the SAI-C model we obtained much lower scores, respectively Precision = 0.727 and Recall = 0.571 for the same confidence threshold. The results indicate a lower classification performance of the SAI-C model compared to the SAI-G model. The exception is the value of Precision for the POSITIVE concept, which is 1,000.
Key words:
Sentiments analysis from images, convolutional neural network.

1. Introduction

Identifying the feelings in the images taken by the video monitoring systems, during a physical or online course, can contribute to the improvement of school performance by correlating with the contents who generated this reactions and emotions.

The observation of these behaviors by a teacher is limited in time, space and as number of course participants. An interesting challenge is to measure the efficiency of an automatic machine that would perform such monitoring activities.

The tens of billions dollars investments in Computer Vision domain have led to the creation of solutions that allow the use of models trained by using impressive amounts of images or the creation of new models that use their own data sets in order to classifying images using abstract labels such as sentiments.

In our study we will use the AutoML Vision service from Google Cloud and the Clarifai platform to create two models, SAI-G (Sentiments Analysis from Images - Google) and SAI-C (Sentiments Analysis from Images - Clarifai) to classify images according to two categories of emotions: positive and negative. Both solutions using convolutional neural networks.

Our goal is to make a comparison between the performances of the two models. For this we will use the evaluation metrics: Precision, Recall, ROC curve (Receiver Operating Characteristic), Precision-Recall curve, Confusion Matrix, Accuracy score and Average precision.

The paper includes the sections: Section 2 who makes a short introduction to computer vision and convolutional neural network, Section 3 describes the working methodology, Section 4 presents the results and Section 5 for conclusions.

2. Theoretical notions

The value of investing in the technologies needed to teach cars to perceive the world as humans do will increase from \$ 2.37 billion in 2017 to \$ 25.32 billion in 2023 [1]. This area of research, named Computer Vision, has applications in the most diverse fields such as: automotive industry, gaming, electronic commerce, security, medicine and others [2].

Solving problems such as self driving or diagnosing pulmonary cancer with an accuracy of about 75%, in the absence of advanced screening procedures, is done using convolutional neural networks [3].

These are deep neural networks in which the hidden layers are the convolutional, pooling and fully connected layers [4] (Fig. 1)

Manuscript received September 5, 2021
Manuscript revised September 20, 2021
https://doi.org/10.22937/IJCSNS.2021.21.9.25

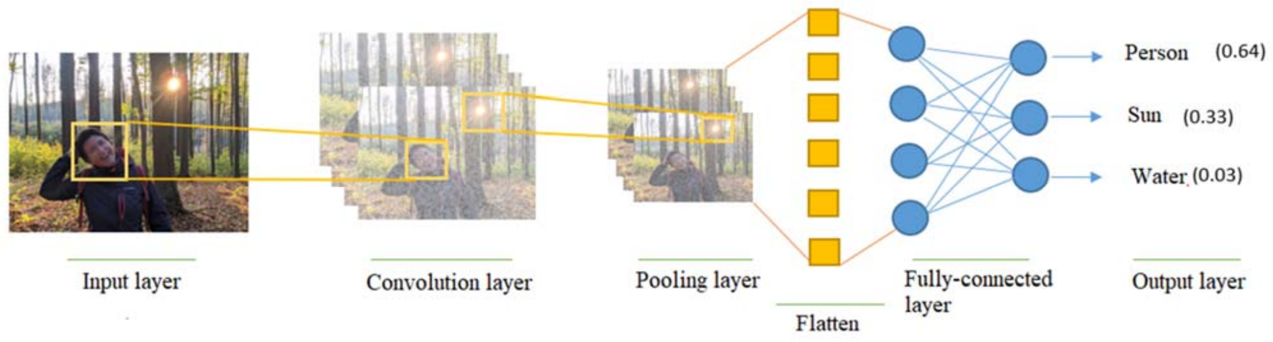

Fig. 1 General architecture of a convolutional neural network.

Each image from the input layer is a 3-dimensional array: the height and width expressed in pixels and the depth given by the number of color channels specific to the encoding used [5].

Convolution is an operation with two functions that describes how the shape of one function is changed by the other function. Mathematically, this operation is described by relation (1). The function f can describe the location of an object, who changes in time, depending by the weighting function g proportional with the time which the measurement takes place [6].

$$(f * g)(t) = \int f(t_0)g(t - t_0)dt_0 \tag{1}$$

From the point of view of convolutional neural networks, convolution means the application over images in the input layer of a filter or kernel, in the form of a multidimensional matrix containing hyperparameters that changes over time until the output' dimensions are an integer (see Eq. 2) [7].

$$O = \frac{I - K + 2Z}{S} + 1 \tag{2}$$

The output size, O, is given by the hyperparameters of the filter that moves over the image with stride S: I - image size, K- filter size, C- number of channels, Z- zero-padding [8].

By activating the movement of the filter over the image can be detected: the orientation of the lines, shadows, color spots that help to detect objects [9].

The number of operations performed by the network in the process of training the model is very large, which is why it is useful to decrease this volume by connecting the pooling layers to the output of the convolution layers. They will compress the volume using the average or maximum value in a region (Fig. 2) [10].

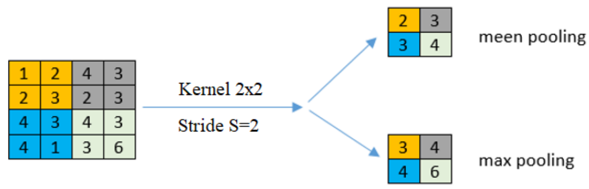

Fig. 2 Pooling operation.

The outputs of the pooling layer are flatten into a vector and then connected to the fully connected layers and then to the output layer.

3. Methodology

3.1 The data set and the working procedure

The data set contains 153 JPG images in which they can be identified: landscapes from various seasons, people, objects, animals, buildings, cakes, handwritten or printed text, books. These were separated into two folders: POSITIVE (76 images) and NEGATIVE (77 images) so that there was a balance between the two categories. These were labeled: positive (POZITIV) or negative (NEGATIV) (Fig. 3).

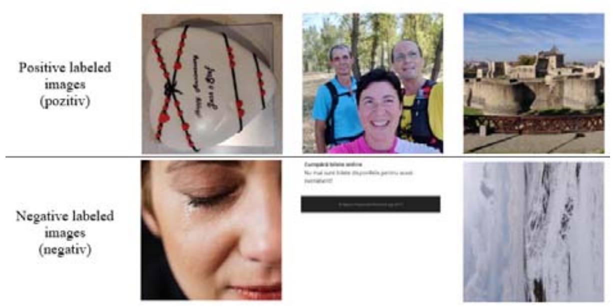

Fig. 3 Labeled images examples.

For the SAI-G model, the data set is divided into the following percentages: 80% for training, 10% for validation and 10% for model testing (Table 1).

Table 1: Data set dividing

Labels	***Train***	***Validation***	***Test***
Negativ (Negative)	64	6	7
Pozitiv (Positive)	58	9	9

The SAI-C model uses 17 Positive (Pozitiv) labeled and 14 Negative (Negativ) labeled images for evaluation.

The AutoML Vision Service from Google Cloud allows supervised training of a model in order to classify the images according to the assigned labels.

The images are uploaded on the Clarifai platform, the Positive (Pozitiv) and Negative (Negativ) concepts are created and then the images are mapped to the created concepts.

3.2 Model evaluation

The performance of the model can be evaluated at the general level or at the level of each category of those labeled with the help of evaluation metrics.

We make the following notations: TP (True Positive) - the number of correctly positive predictions, TN (True Negative) - the number of correctly negative predictions, FP (False Positive) - the number of false positive predictions and FN (False Negative) - the number of false negative predictions. According to these notations, Precision and Recall are defined by the following relations (See Eqs. 3, 4) [11]:

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

We will use the confidence threshold, with values in [0, 1], to adjust the Precision and Recall. A graphical representation of these changes is called the Precision-Recall curve [12]. The area under the Precision-Recall curve is given by the relation (See Eq. 5), in which P(R) is the function that defines the modification of Precision and Recall for different values of the confidence threshold [13].

$$AUC = \int_0^1 P(R)dR \tag{5}$$

4. Results

4.1 Results obtained with the SAI-G model

Average precision is a metric that measures the performance of the model calculated as the area under the Precision-Recall curve, so a measure of the change of Precision and Recall depending on the evolution of the confidence threshold in the range [0,1] [14].

The SAI-G model obtains an equal score for Precision and Recall, namely 87.5% at a value of the confidence threshold of 0.5, and for Average Precision a value of 0.929. At the level of each category the results were obtained: for the Positive class, Precision and Recall are equal to 88.89%, and for the Negative class, Precision and Recall have the value of 85.71%. These are calculated for a confidence threshold value of 0.5.

These values can be visualized using the confusion matrix (Fig. 4) in which we also observe that 11% of the images labeled Positive (Pozitiv) were classified as Negative (Negativ), while 14% of the images labeled Negative were assigned to the other category.

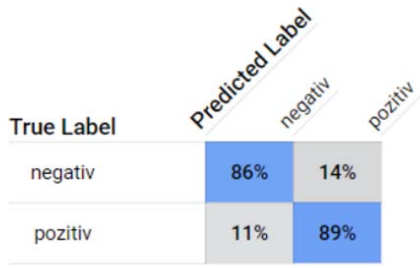

Fig. 4 Confusion matrix.

Figure 5 shows the Precision-Recall curves at the general level and at the level of each category.

With the change of the confidence threshold, Precision and Recall tend to keep their values constant (Fig.6), which means that the model is stable in connection with predictions for the two labeled values.

Some examples of images classified correctly in relation to the assigned labels are represented in figure 7 (for the Positive value) and figure 8 (for the Negative value).

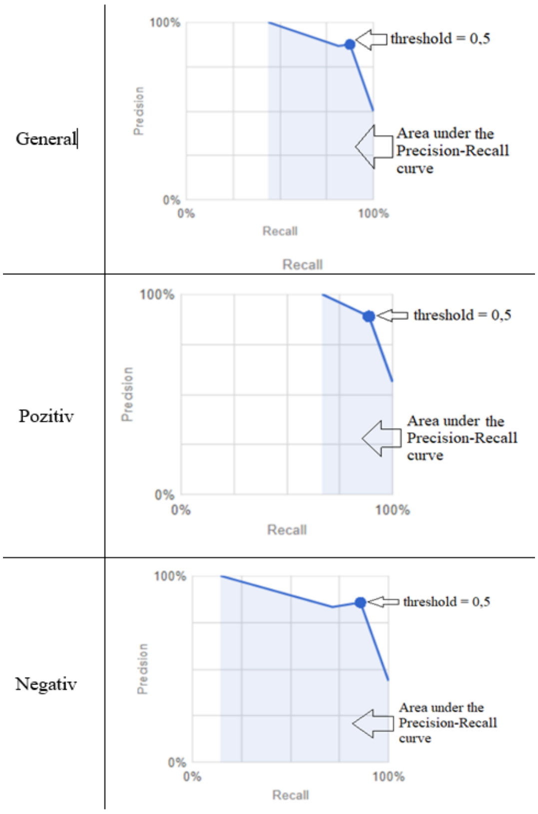

Fig. 5 Precision-Recall curves.

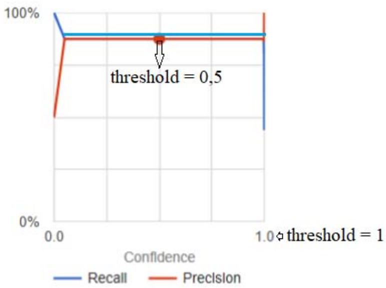

Fig. 6 Changing the Precision and Recall with increasing confidence threshold.

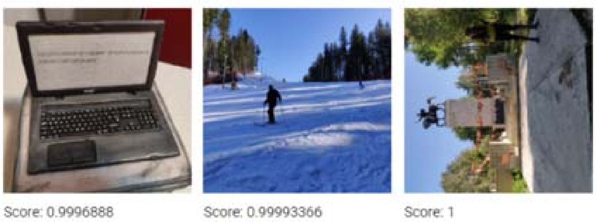

Fig. 7 Correctly classified images with Positive label.

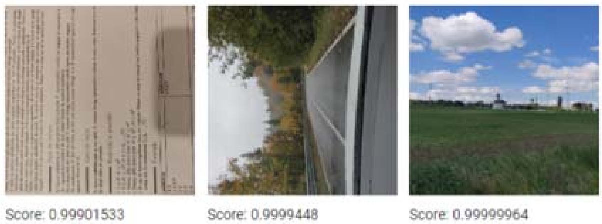

Fig. 8 Correctly classified images with Negative label.

4.2 Results obtained with the SAI-C model

The ROC (Receiver Operating Characteristic) curve or receiver operating characteristic curve is a planar representation of the quantities TPR (OY axis) and FPR (OX axis), defined by relations 6 and 7 (See Eqs. 6, 7) [15].

$$TPR = \frac{TP}{TP + FN} \tag{6}$$

$$FPR = \frac{FP}{FP + TN} \tag{7}$$

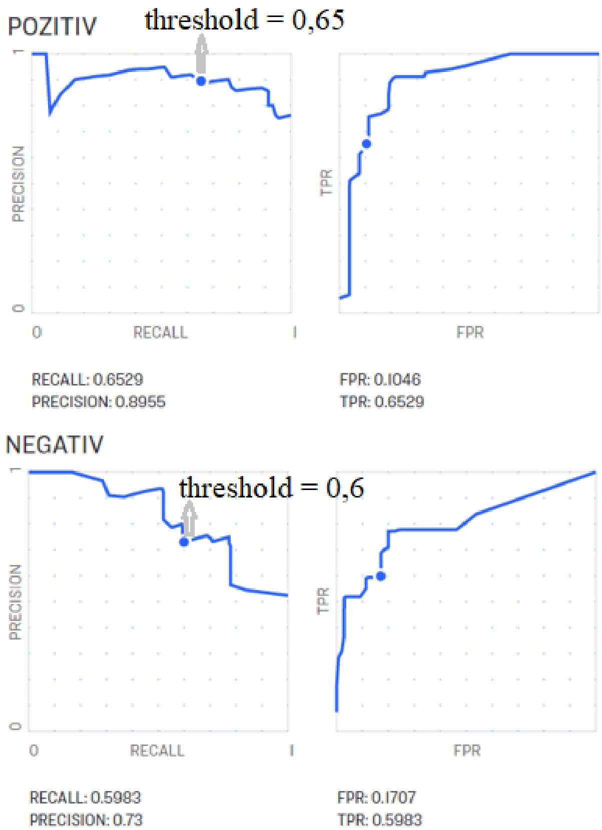

Fig. 9 Precision-Recall and ROC curves for the concepts POZITIV (Positive), NEGATIV (Negative).

If threshold value is fixed at 0,65, the Precision is 0,6529 and the Recall is 0,8955 for the POZITIV concept. For the NEGATIV concept, a confidence threshold score of 0,6 produces the values of 0,5983 for Precision and 0,73 for Recall (Precision-Recall curves, Fig. 9).

TPR values are approximately the same for the FPR values in the range [0, 0.1] (ROC curve for NEGATIV, POZITIV concepts in Fig. 9).

The results obtained by the SAI-C model that uses the Clarifai platform, at the general and at the each of the two concepts level, POZITIV (Positive) and NEGATIV (Negative), are presented in Table 2.

Table 2: The results obtain by the SAI-C model

Concept	*Accuracy*	*Recall*	*Precision*
NEGATIV	0,811	0,571	0,727
POZITIV	0,888	0,647	1,000
General	0,849	0,609	0,864

4.3 Comparison between the results obtained with the SAI-G and SAI-C models

Table 3 presents comparative results between the model created with the Google AutoML Vision service and the model created on the Clarifai platform in terms of the values of classification accuracy in general, Precision and Recall in general and for each of the two categories: POZITIV, NEGATIV.

Table 3: Comparison between the SAI-G and SAI-C models

Evaluation metrics	*SAI-G Model*	*SAI-C Model*
Accuracy	0,929	0,849
Precision	0,875	0,609
Recall	0,875	0,864
Precision at the POZITIV class level	0,888	1
Recall at the POZITIV class level	0,888	0,647
Precision at the NEGATIV class level	0,857	0,727
Recall at the NEGATIV class level	0,857	0,571

5. Conclusions

In our paper, we built two models using two specific services: AutoML Vision from Google and the Clarifai platform.

The SAI-G and SAI-C models classify 153 labeled images into two categories: positive (POZITIV) and negative (NEGATIV). The data sets used for the two labeled concepts are balanced, 76 images being mapped to the POZITIV and 77 to the other category. 10% of these images are used by AutoML Vision to evaluate the model, while Clarifai uses 20.26% of the images for this purpose.

The SAI-G model obtains, at the level of both categories, equal scores for Precision and Recall (87.5%), calculated at a confidence threshold of 0.5 and an average precision of 0.929. At the level of each category, this model obtains the same score for Precision and Recall, for each of the labeled values. We find that the Precision of classifying images in the POZITIV category is higher than that of classifying images in the NEGATIV class, namely 88.89% compared to 85.71%, values calculated at the same confidence threshold of 0.5.

By comparison, the SAI-C model obtains lower scores for Precision and Recall in general level and at the level of each concept, except for Precision obtained in the classification of images mapped to the POZITIV concept which has the value 1,000.

In the case of the SAI-G model, we observe from the Precision-Recall curve that an evolution of the confidence score in the range [0, 1] does not produce significant changes of Precision and Recall. This means that the classifications made are stable for both categories of labeled images in relation to the change in the number of True Positive and False Negative cases.

Identifying the emotions in the images captured by the monitoring devices during the educational processes and correlating them with the learning contents is a new and important research direction that can have remarkable results in terms of increasing learning performance.

Acknowledgments

„This work is supported by the project ANTREPRENORDOC, in the framework of Human Resources Development Operational Programme 2014-2020, financed from the European Social Fund under the contract number 36355/23.05.2019 HRD OP /380/6/13 – SMIS Code: 123847”.

References

[1] Market and Markets, "*AI in Computer Vision Market by Component (Hardware, Software), Vertical (Automotive, Sports & Entertainment, Consumer, Robotics & Machine Vision, Healthcare, Security & Surveillance, Agriculture), and Region - Global Forecast to 2023*". Online: https://www.marketsandmarkets.com/Market-Reports/ai-in-computer-vision-market-141658064.html

[2] Mark J. P. Anson; Frank J. Fabozzi; Frank J. Jones (3 December 2010). *The Handbook of Traditional and Alternative Investment Vehicles: Investment Characteristics and Strategies*. John Wiley & Sons. pp. 489–. ISBN 978-1-118-00869-0 (2010)

[3] W. Ausawalaithong, A. Thirach, S. Marukatat and T. Wilaiprasitporn, "*Automatic Lung Cancer Prediction from Chest X-ray Images Using the Deep Learning Approach*," 2018 11th Biomedical Engineering International Conference (BMEiCON), Chiang Mai, 2018, pp. 1-5, doi: 10.1109/BMEiCON.2018.8609997 (2018)

[4] Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., ... & Chen, T. (2018). *Recent advances in convolutional neural networks*. Pattern Recognition, 77, 354-377 (2018).

[5] O'Shea, Keiron & Nash, Ryan. (2015). *An Introduction to Convolutional Neural Networks*. ArXiv e-prints (2015).

[6] Ian Goodfellow and Yoshua Bengio and Aaron Courville (2016). *Deep Learning*. MIT Press. p. 326, chapter 9. Online: https://www.deeplearningbook.org/contents/convnets.html. (2016)

[7] Afshine Amidi, Shervine Amidi, "*Super VIP Cheatsheet: Deep Learning*" (2018), Stanford University, Chapter 1, pp: 2-7. (2018)

[8] S. Albawi, T. A. Mohammed and S. Al-Zawi, "*Understanding of a convolutional neural network*," 2017 International Conference on Engineering and Technology (ICET), Antalya, 2017, pp. 1-6, doi: 10.1109/ICEngTechnol.2017.8308186 (2017)

[9] Lee, Honglak & Grosse, Roger & Ranganath, Rajesh & Ng, Andrew. (2009). *Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations*. Proceedings of the 26th International Conference On Machine Learning, ICML 2009. 77. 10.1145/1553374.1553453 (2009)

[10] ROVISLaboratory, *Machine Learning*, Curs 8 – Convolutional Neural Network. Online: https://rovislab.com/courses/ml/08_NeuralNetworks-CNN/08_Retele_Neurale_Convolutionale.pdf

[11] Scikit-Learn documentation, "*Precision-Recall*". Online: https://scikit-learn.org/stable/auto_examples/model_selection/plot_precision_recall.html

[12] Keilwagen, Jens & Grosse, Ivo & Grau, Jan. (2014). *Area under Precision-Recall Curves for Weighted and Unweighted Data*. PloS one. 9. e92209. 10.1371/journal.pone.0092209 (2014)

[13] Boyd, Kendrick & Eng, Kevin & Page, C.. (2013). *Area under the Precision-Recall Curve: Point Estimates and Confidence Intervals*. 8190. 451-466. 10.1007/978-3-642-40994-3_29 (2013)

[14] Zhang E., Zhang Y. (2009) Average Precision. In: LIU L., ÖZSU M.T. (eds) Encyclopedia of Database Systems. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-39940-9_482 (2009)

[15] Kumar, Rajeev & Indrayan, Abhaya. (2011). *Receiver Operating Characteristic (ROC) Curve for Medical Researchers*. Indian pediatrics. 48. 277-87. 10.1007/s13312-011-0055-4. (2011)

Daniela Marcu received her BSc in Physics (1997), postgraduate studies in Computer Science (2001), PhD student in Computers and Information Technology. Now she is professor of computer science at the Stefan cel Mare College. She is the author/co-author of 18 books and some papers which have been published in journals and presented at different conferences. Her current research interests include different aspects of Big Data applied in the educational field, Natural Language Processing, Computer Vision, Sentiment Analysis, Algorithms Programming in C/C++ Language. Personal Website: https://modinfo.ro/.

Mirela Danubianu has obtained the B.S. and M.S. degree in Computer Science from University of Craiova in 1985, and the PhD. degree in Computer Science in 2006 from Stefan cel Mare University of Suceava. She has also obtained the B.E. degree in Economics from University of Craiova in 2001. Currently, she is Associate Professor and Head of the Computers Department at Stefan cel Mare University of Suceava. She is the author/co-author of 5 books, 7 chapters and more than 100 papers which have been published in journals and presented at different conferences. Her current research interests include databases theory and implementation, modern data architectures, data analytics, application of Data Science in economics, education and healthcare.