

How Artificial Intelligence impacts the programs broadcast by Globo TV: Case studies

Alvaro Antelo
Edmundo Hoyle

CITE THIS ARTICLE:

Antelo, Alvaro and Hoyle, Edmundo; 2019. How Artificial Intelligence impacts the programs broadcast by Globo TV: Case studies. SET INTERNATIONAL JOURNAL OF BROADCAST ENGINEERING. ISSN Print: 2446-9246 ISSN Online: 2446-9432. doi: 10.18580/setijbe.2019.3. Web Link: <http://dx.doi.org/10.18580/setijbe.2019.3>



COPYRIGHT This work is made available under the Creative Commons - 4.0 International License. Reproduction in whole or in part is permitted provided the source is acknowledged.

How Artificial Intelligence impacts the programs broadcast by Globo TV: Case studies

Alvaro Antelo, Edmundo Hoyle

Abstract—In this article we show how Artificial Intelligence (AI) are impacting multimedia content creation, processing and delivery in the programs broadcasted by Globo TV. AI is mostly perceived as a new tool capable of optimizing and enhancing processes through automation in ways that were unthinkable in the recent past. Three broadcast programs were chosen as targets for AI algorithms deployment in our current workflow aiming at increased productivity, cost reduction and novel ways to generate content.

Index Terms—Artificial Intelligence Deep Learning Image Processing Broadcast TV.

I. INTRODUCTION

ADVANCES in AI have enabled manufacturers and broadcasters to develop and implement smarter tools and applications to boost the entire content lifecycle.

Deep learning is an aspect of AI that is concerned with emulating the learning approach that human beings use to gain certain types of knowledge. The phrase “deep learning” probably conjures up images of sentient robots staging a hostile takeover. But in reality, deep learning is just another way to describe large neural networks, a technology you encounter every day when you browse the Internet or use your mobile phone. A basic neural network might have one or two hidden layers, while a deep learning network might have dozens or even hundreds.

Today many AI applications exist around us, and we use them normally without perceiving it. For example when somebody is looking for images that included an object of interest. A short list with the main applications could be similar to this:

- Automatic Machine Translation.
- Object Classification in Photographs.
- Automatic Handwriting Generation.
- Character Text Generation.
- Self-driving vehicles
- Image Caption Generation.
- Automatic object segmentation.
- Colorization of Black and White Images.

In this article we explain how Globo TV utilizes common AI techniques to increase its productivity, reduce costs and create novel ways to generate content. The organization of this article is as follows: The first section shows a brief summary about the three techniques of deep learning used by Globo TV in each of their programs broadcasted. In the second section

we’ll talk about how each broadcast program was impacted by the techniques of deep learning described in the previous section. The workflow used in the process to exploit the AI algorithms will be showed in the fourth section. The last section contains the conclusions about our experiences.

II. STATE-OF-ART

In this section we describe briefly the three techniques of deep learning that were used in three programs broadcasted by Globo TV. Starting with neural style transfer which was used in *The Voice kids*, followed by automatic colorization used in *Cidade Proibida* and ending with semantic segmentation employed in segmentation the “secret reporter” in the program *Fantástico*.

A. Neural Style transfer

In 2015, Gatys et al. [1] showed a seminal work that demonstrated how convolutional neural networks (CNN) could be used to generate stylized images from a texture example. This process of using a CNN to render a content image in different styles is referred to as Neural Style Transfer. Their experimental results demonstrated that the content and style in a photo were separable, which indicates the probability of changing a photo’s style while preserving desired semantic content as showed in Figure 1.

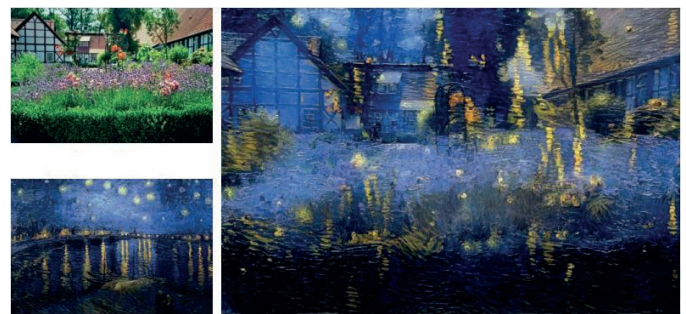


Fig. 1: A simple sample using Neural Style Transfer. Upper left: Input image, lower left: style, right side: output image

A wide range of implementations have been made freely available in [2] [3] [4], based varyingly on different neural network frameworks such as Caffe [5] and Tensorflow [6] and wrappers such as Torch and PyCaffe [7]. Joshi et al. in [8] shows a high-level case study about what it’s like to apply style transfer techniques to cinema, they used the technique to redraw key scenes in the short movie *Come Swim* in the style of the impressionistic painting that inspired the film. Figure 2 shows a sample from the movie.

A. Antelo and E. Hoyle are with Department of Research and Development of TV Globo.

E-mail: alvinho, edmundo.hoyle @tvglob.com.br



Fig. 2: Usage of Neural Style Transfer in Come Swim; top: content image, top right : style image, bottom: upsampled result. Images taken from [8]

B. Automatic Colorization

The power of convolutional neural networks for image processing is mostly demonstrated on the seemingly impossible task such as colorizing a black and white picture. Anyone familiar with color conversion math will readily attest that once color information is lost there is no way to restore it. Deep learning proved that it is possible to colorize a black and white picture because as Zhang *et al.* said [9] “in many cases, the semantics of the scene and its surface texture provide ample cues for many regions in each image: the grass is typically green, the sky is typically blue, and the ladybug is most definitely red”; also citing Larsson [10] “Automatic colorization serves as a proxy measure for visual understanding”, implying a close relation between colorization and semantic segmentation; in other words the network must be able to detect and segment visual objects prior to assigning a probable color to its region. In fact, most convolutional networks trained for colorization begin with networks trained for object detection and classification and are further tuned for colorization with a large dataset of color images. Computerized colorization techniques existed prior to machine learning but they relied on human artists to scribble color on image regions and a computer assisted the colorization process by completing the color fill throughout similar regions. The first fully automated colorization pipeline goes back to Cheng *et al.* [11] in which various features are extracted from the images and the different patches of the image are colorized using a small neural network. Automatic colorization using deep



Fig. 3: Black and white images colorized using automatic colorization. Images taken from [12]

learning predicts plausible colors to image regions without any human intervention. To accomplish this, a neural network has to be trained with millions of reference color images, exposed to only the luminance channel and as it tries to estimate the colors, any deviations from ground truth on the pixel level at the output gets backpropagated inside the network in the form of error information and the whole process iterates as it converges during several thousands iterations throughout all training images. We found that the most common colorspace used in research is CIE LAB because it is designed to be a perceptually linear space but it is converted to YUV or RGB as needed on the output. During the experiments, we found that the presence of significant amounts of noise and also low contrast images on the black and white pictures reduced the colorization performance of all networks, so we introduced a pre-processing stage which does noise reduction and contrast normalization prior to the colorization.

C. Semantic segmentation

Semantic segmentation is the task of clustering parts of images together which belong to the same object class. For this to happen it is necessary to identify the classes of objects in images and also to localize them inside the images with reasonable accuracy. The segmentation of objects is one of the key problems in the field of computer vision, the Figure 4 explains the evolution of object segmentation. Garcia-Garcia *et al.* said in [13] “Looking at the big picture, semantic segmentation is one of the high-level task that paves the way towards complete scene understanding”. Many applications depend on accurate semantic segmentation such as autonomous driving, human-machine interaction, image search engines and augmented reality. It is important to distinguish semantic segmentation from object detection which only has to distinguish different instances of the same object in the image; for example the image of a car behind a tree is split into two objects using object detection but only one car with semantic segmentation. There exists a wide range of segmentation algorithms, mainly

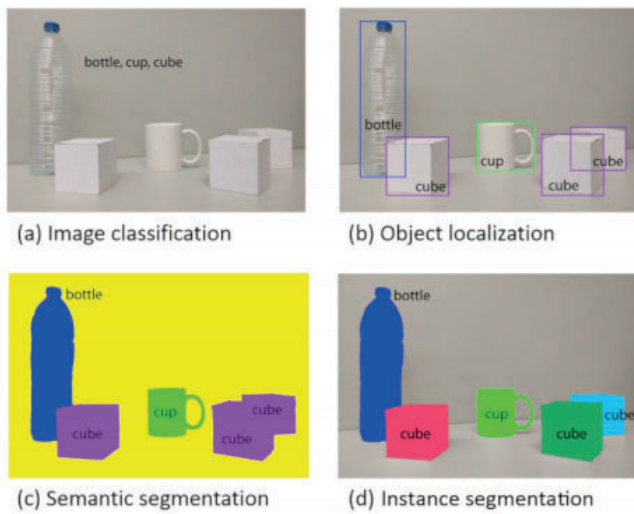


Fig. 4: Evolution of object recognition or scene understanding from coarse-grained to fine-grained inference: classification, detection or localization, semantic segmentation, and instance segmentation. Image taken from [13]

varying on the number of classes of objects it's able to process, for example an autonomous car system has a semantic segmentation based on binary classes such as street versus non-street or more classes as road signs, people, vehicles, etc. Deep neural networks, usually convolutional neural networks dominate this field nowadays and their accuracy far surpasses any other methods of segmentation. In order to train a neural network for this task, one has to start with a dataset of images previously annotated, in other words the images must be labeled with masks identifying all classes that will have to be learned, for example the image of a car has to have all pixels containing the car marked as a single region, if there is a person in that image too it has to be labeled as a different class. Google provided the Tensorflow Object Detection API [14], this API made the segmentation objects such as persons, cars and many others as shown in Figure 5.

III. CASE STUDIES

This section describe how the techniques showed in the previous section were used in the GLOBO TV programs.

A. The Voice Kids

The Voice Kids is a version of The Voice tv series franchise in which kids participate. The Brazilian version went on air on 2016, and in this program, singers perform on a stage with big screens on the background. The technique was used to create the videos that go on the big screens behind the participants of The Voice Kids. The neural style transfer was our first initiative to produce content using Artificial Intelligence methods on Globo TV. In this program the participant Juan Carlos Poca was going to sing the song “o amor” (Zezé Di Camargo and Luciano) and the team needed soft videos that could keep pace with the music. They had already thought of using landscapes and adding some kind of effect on them. Various



Fig. 5: Segmentation and recognition sample using the Tensorflow Object Detection API

styles and paintings of artists were trained and applied directly over videos select by the producers, and the paintings of impressionist artist Claude Monet were selected as the style that produce the best results. Figure 6 shows two image samples extracted from the videos showed in The Voice Kids Brasil, and in [15] it is possible to view the performance. This is a novel way to generate content with relative low cost.



Fig. 6: Images from the Voice Kids Brasil with the screens showing images created with neural style transfer

B. Cidade Proibida

Cidade Proibida "Forbidden city" is a twelve episodes Globo TV series aired weekly beginning on September 2017. It's a detective story set on 1950 Rio de Janeiro and uses many old black and white film footage. The clips were used in the context of flashbacks of the main character and as short sonorized clips exhibited whenever the main character moves around the city. It was decided that these clips would be colorized in order to appeal more to the audiences. Image colorization bureaus such as Dynamichrome [16] company do this on a manual basis with superb color rendition but costs can go up to US\$ 100,00 per image and time spent depending on the image can take from a day to an entire month. Total duration of footage per episode was around fifteen seconds which at 24 frames per second gives 360 frames per episode, or 4320 frames for the whole series. This amount of frames makes it impossible to tackle with colorization on a manual basis, because of the high cost and time needed to colorize all frames. In order to decrease the costs we looked at automatic colorization using deep learning as a valid option. In our tests the processing time to colorize was two seconds per frame, therefore the process to colorize all frames was 144 minutes. We used custom linux images on g3.4xlarge virtual instances in Amazon EC2 to provide us with the necessary computing power to achieve the result in a reasonable amount of time. The machine cost is \$ 1,14 per hour, totaling approximately \$ 200,00 to colorize all videos. Figure 7 show two sample frames with theirs respective images colorized that were used in the series *Cidade Proibida*

C. Fantástico

Fantástico is a Brazilian weekly television news magazine broadcast on Sunday nights on Globo TV since August 5, 1973. It is an open-ended newscast with hosts introducing and reporting stories. One of the segments, named "Secret Reporter" is hosted by Eduardo Faustini, a brazilian journalist, being one of the most important investigative reporters of Globo TV. To preserve his physical integrity, he never revealed his identity in these investigations, so there is a need to obfuscate his image in post production. For each story, the time spent in hiding the reporter was taking about 3 days. This required the team to meet a week in advance, each time a new story will be produced to determine an upper limit on the amount of material that could be used. It is a painstaking process that takes days to create only a few seconds of video. The main objective of this project was to reduce the time spent by the art team in hiding the reporter in the video scenes aired in *Fantástico*. Deep learning semantic segmentation was employed to identify, localize and masquerade Faustini's image in an automated fashion. Semantic segmentation is implemented in a deep learning network that was trained with thousands of images segmented and manually annotated, in these images there are people, cars, chairs, dogs etc. With the trained network this algorithm returns the masks of the objects found in the image, be it people, cars or any other object found. In our case we only used the segmentation marked as a person. After the processing, the algorithm returns a video containing



Fig. 7: Original Black and white frames and their respective colorization used in *Cidade Proibida*

the masking area to be refined by the Arts team. The whole process takes about 1 minute for each second of video, and due to the fact that it can be run in parallel, now it makes no difference if 1 or 20 videos arrive at the same time. With the reduction in processing time, it is now possible to receive a greater amount of material and it opens the possibility of



Fig. 8: The Secret Reporter hidden behind the blue mask

including new material received at the last-minute. In addition to the gains in productivity, the new process allows the artist to be more focused on his main work (doing art) and not on a painstaking manual process of creating masks. The cost to keep the system running on Amazon is around 50 dollars per month, a very low cost for the huge benefits brought to the production of the program. Figure 8 show a few images showing the Secret Reporter hidden behind the blue mask. There is also a link for a video showing some footage from the program [17]

IV. WORKFLOW

All cases described in this paper use their own deep learning technique but the process life cycle is very similar between them. The infrastructure for this process was built on Amazon Web Services (AWS) in order to decouple them from our internal workflow and also to leverage elastic scaling, in other words the infrastructure would be deployed on demand as

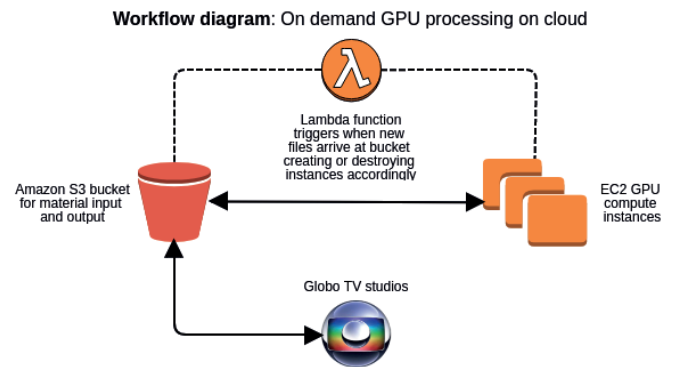


Fig. 9: AWS Workflow process

new videos got uploaded and destroyed as soon as it is not necessary anymore. In order to store the video files before and after processing we used Amazon S3 simple storage service which is an object storage accessed by web services interfaces; in practice an editor would open a web interface and upload video clips to be processed. As soon as the files got uploaded, AWS lambda which is an event driven serverless computing platform would trigger the instantiation of GPU machines containing the AI algorithms, ready to begin consuming the newly received video files. For each video file a new machine spun up so processing occurred on a parallel fashion. At the end of processing, the resulting video files got written on the S3 bucket and the machine which finishes the job got destroyed. The editor can therefore look at the output bucket on S3 and start downloading the processed video files. Figure 9 shows the workflow process in AWS.

V. CONCLUSION AND FUTURE EXPLORATIONS

In this paper we presented three applications of deep learning algorithms applied on the production of three broadcast programs. In all cases we saw benefits be it on the productivity aspect or in novel ways to generate content. Using cloud deployment makes it easy to just plug the algorithms in our current workflow minimizing eventual integration problems. Artificial intelligence opens up new opportunities for productivity enhancement in the broadcast environment.

Looking from the sole perspective of automation, we see many avenues for innovation in video related workflows, especially in human related tasks closely tied to visual understanding, for example tagging images, searching for images, archiving material, in quality control and error detection on video, switching video based on content, etc. Those tasks are currently assigned to humans but can be significantly offloaded to AI algorithms providing massive improvements in throughput and possibly error reduction.

REFERENCES

- [1] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. A neural algorithm of artistic style. *CoRR*, abs/1508.06576, 2015.
- [2] Frank Liu. style-transfer. <https://github.com/fzliu/style-transfer>. Accessed: 2018-06-30.
- [3] Justin Johnson. neural-style. <https://github.com/jcjohnson/neural-style>. Accessed: 2018-06-30.

- [4] Anish Athalye. neural-style. <https://github.com/anishathalye/neural-style>. Accessed: 2018-06-30.
- [5] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross B. Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. *CoRR*, abs/1408.5093, 2014.
- [6] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Gregory S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian J. Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Józefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Gordon Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul A. Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda B. Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *CoRR*, abs/1603.04467, 2016.
- [7] Soheil Bahrampour, Naveen Ramakrishnan, Lukas Schott, and Mohak Shah. Comparative study of caffe, neon, theano, and torch for deep learning. *CoRR*, abs/1511.06435, 2015.
- [8] Bhautik Joshi, Kristen Stewart, and David Shapiro. Bringing impressionism to life with neural style transfer in come swim. In *Proceedings of the ACM SIGGRAPH Digital Production Symposium*, DigiPro '17, pages 5:1–5:5, New York, NY, USA, 2017. ACM.
- [9] Richard Zhang, Phillip Isola, and Alexei A. Efros. Colorful image colorization. *CoRR*, abs/1603.08511, 2016.
- [10] Gustav Larsson, Michael Maire, and Gregory Shakhnarovich. Learning representations for automatic colorization. *CoRR*, abs/1603.06668, 2016.
- [11] Zezhou Cheng, Qingxiong Yang, and Bin Sheng. Deep colorization. *CoRR*, abs/1605.00075, 2016.
- [12] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. Let there be Color!: Joint End-to-end Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification. *ACM Transactions on Graphics (Proc. of SIGGRAPH 2016)*, 35(4):110:1–110:11, 2016.
- [13] Alberto Garcia-Garcia, Sergio Orts-Escolano, Sergiu Oprea, Victor Villena-Martinez, and José García Rodríguez. A review on deep learning techniques applied to semantic segmentation. *CoRR*, abs/1704.06857, 2017.
- [14] Jonathan Huang, Vivek Rathod, Chen Sun, Menglong Zhu, Anoop Korattikara, Alireza Fathi, Ian Fischer, Zbigniew Wojna, Yang Song, Sergio Guadarrama, and Kevin Murphy. Speed/accuracy trade-offs for modern convolutional object detectors. *CoRR*, abs/1611.10012, 2016.
- [15] The Voice Kids Brazil. Juan carlos poca performance. https://www.youtube.com/watch?v=6w_FM47iRp0. Accessed: 2018-06-30.
- [16] Dynamichrome. <http://dynamichrome.com/>.
- [17] Fantantisco. Reportersecreto. <https://globoplay.globo.com/v/6716012/programa/?s=3m31s>. Accessed: 2018-06-30.



Edmundo Hoyle received his Bsc in Physics at National University of Trujillo (Peru) in 2004 and his DSc at the Federal University of Rio de Janeiro (Brasil) with specialization in Image Processing in 2013. Actually he works as researcher at TV Globo and his main research interests are image processing, computer vision and deep learning.



Alvaro Antelo received his BSc at the University Veiga de Almeida of Rio de Janeiro (Brasil) in 2001. Actually he works as researcher at TV Globo and his main research interests include deep learning, video over IP and networking.

Received in 2018-06-30 | Approved in 2019-07-10