Autonomous identification of high-contact surfaces from convolutional neural networks

Angie Alonso¹, Andres Peña², and Fredy Martínez³
Universidad Distrital Francisco José de Caldas, Facultad Tecnológica, Cra. 7 No.40B-53, Bogotá D.C., Colombia
E-mail: ¹analonsol@correo.udistrital.edu.co, ²anfpenar@correo.udistrital.edu.co, ³fhmartinezs@udistrital.edu.co

Abstract. The rapid spread of the SARS-CoV-2 virus has highlighted many social interaction problems that favor the spread of disease, particularly airborne spread, which can be addressed by adjusting existing systems. Of particular interest are places where large numbers of people interact, as they become a focus for the spread of these diseases. This paper proposes and evaluates an autonomous identification scheme for certain surfaces considered high risk due to their continuous handling. These high-contact surfaces can be identified by an autonomous system to apply specific cleaning tasks to them. We evaluate three convolutional models from a proprietary dataset with a total of 2000 images ranging from wall switches to water dispensers. The objective is to identify the ideal architecture for the system. The ResNet (Residual Neural Network), DenseNet (Dense Convolutional Network), and NASNet (Neural Architecture Search Network) models were selected due to their high performance reported in the literature. The models are evaluated with specialized metrics in non-binary classification problems, and the best scheme is selected for prototype development.

1. Introduction
Contamination of high-contact surfaces has now become a focus of special attention to minimize the spread of viruses, including SARS-CoV-2 and others such as enteric viruses [1]. These viruses can remain on surfaces for several weeks, or even months, and act as sources of contagion and potential spread [2]. The surfaces at risk are classified as porous and non-porous, the latter referring for example to stainless steel and plastics, among others, and are surfaces on which it has been proven that viruses can remain for a longer time [3]. The permanence of viruses on these surfaces depends on the frequency with which they are cleaned, a strategy that has been implemented for the benefit of public health is the disinfection of surfaces more routinely, and increased hand hygiene practices to reduce viral transmission in any space with high interaction of people [4].

One of the main sectors affected by the current pandemic has been the tourism sector [5]. Closures, quarantines, and mobility controls have considerably reduced the economic movement of this sector [6]. For example, the World Tourism Organization (WTO) estimates a worldwide drop in tourism of between 20% and 30% [7, 8]. For this reason, the hotel sector is in the task with implementing strategies to encourage visitors to return to their facilities safely. In this context, the care and maintenance of high-touch surfaces are of vital importance, both to provide safety and to prevent them from becoming a source of contagion [9].
The adopted strategy of this research contemplates the training of a series of classification models through the use of artificial neural networks (ANN) and their possible application in the processing of images related to this type of surfaces. Artificial neural networks date back to 1943 with the first neural model presented by McCulloch and Walter Pitts, and from which the starting point for the development of current deep neural models emerged [10, 11]. The application area we will focus on is image processing, a process that includes several levels. First of all, there is a pre-processing where the information of an image is adapted, and better profiling of the parameters is to be analyzed and identified [12]. For example, brightness and contrast, which can hide or highlight important features in the images. Secondly, segmentation, where the image is partitioned into several regions to better represent the information in detail [13]. The third is the detection and classification of the objects found in the image, and finally the analysis of the image where we obtain the most information of what the image shows [14].

Our objective is to find a high-performance convolutional model that can be integrated into an artificial system with digital cameras, to guide maintenance processes in high contact surfaces [15]. In this sense, it is pertinent that the data obtained by the neural network correspond to the characteristics of the problem, which involves the real-time operation and high capacity of identification in real human environments. Neural networks are capable of identifying specific common parameters in the input data and allow obtaining a closed system that produces output without performing complex mathematical calculations. In this sense, it is convenient to use them in a low-cost embedded system that does not require high processing and/or memory requirements, and that allows fast-updating according to supervised training. Similar schemes have been used to support the fight against SARS-CoV-2, most of them in the development of embedded systems that allow the diagnosis of the disease, but they can be used in other contexts such as the one proposed here [16–18].

Our problem is focused on spaces related to tourist environments, in general, indoor spaces characteristic of hotels and similar. To solve the problem we will describe the methodology to be used both to create the database and to train the different convolutional models. These models will be evaluated using performance metrics to identify the most suitable architecture for use in an embedded prototype [19–21]. In the final part, the results of each model are discussed, and schemes for model improvement and further research are proposed.

2. Problem statement
The tourism sector has grown exponentially in recent years and has become a reference in the world economy. However, nowadays, and with the crisis that arose due to the SARS-CoV-2 pandemic, this sector was faced with a generalized crisis, and its vulnerability to effects that directly affect it became evident. Studies have shown that high-touch surfaces such as door handles, light switches, or laptop computers can spread the virus more rapidly to other surfaces, and people, if they are not properly sanitized. For this reason, it is important to implement strategies to make economic activity compatible with the need to contain the virus and avoid outbreaks.

The objective of this research is to go a step beyond traditional cleaning services and to be able to offer a disinfection service that allows, hand in hand with technology and research on neural networks, its use in real environments with a high turnover of people. The aim is to develop an image-based categorization model capable of autonomously identifying critical elements for a cleaning process in hotels and equivalent spaces. This strategy seeks to regain the confidence of tourists when they return to these vacation destinations, providing them with assurance about the cleaning policies. The methodology proposed to address the problem is shown in the flow chart in Figure 1.

The recognition system must be able to learn the parameters of interest from a proprietary dataset developed by the research group for this purpose. This means that the model must
be able to identify specific parameters in the images regardless of their location, additional elements in the image, size, or orientation. The evaluation of the classification model is performed using the metrics Accuracy, Precision, Recall, F1-score, and confusion matrix. In addition, this evaluation must consider both average performance and performance by category.

3. Methods
By design, convolutional neural networks correspond to receptive fields of neurons in the visual cortex of a biological brain. Consequently, it is assumed that they can identify parameters in images in a similar way as humans do. Therefore, a convolutional model must be trained with a large set of images containing the features that make them different, in the same way that humans learn to identify elements throughout their lives.

Among the different existing convolutional models, we have selected three for their previous high performance on similar problems. We selected the ResNet (Residual Neural Network), DenseNet (Dense Convolutional Network), and NASNet (Neural Architecture Search Network) architectures to evaluate their performance with our dataset, and select the most suitable one to implement our autonomous identification system. The ResNet type networks are characterized by implementing blocks with information jumps forward in the network, which has been observed to increase the capacity of the system. Following the same idea, the DenseNet architecture performs these same jumps forward but in a denser way than the ResNet model, and instead of adding the information it concatenates it, achieving an important reduction in parameters. One of the main improvements of DenseNet is that it manages to shorten the connections between layers near the input and output, thus increasing the density of the network. Finally, the NASNet architecture is characterized by blocks optimized for certain databases, which are then generalized to the network and other similar classification problems. The three selected
models have the possibility of using reduced-size models that would eventually facilitate their implementation on an embedded system.

Our dataset is composed of 20 categories, which include elements typically found in hotels such as wall switches, wall sockets, handrails, door handles, window, and drawer handles, among others. Each category had a total of 100 images, some downloaded from public repositories, and others captured in the laboratory. Figure 2 shows an example of one of the categories of our dataset. All images were subjected to preprocessing consisting of channel adjustment, contrast enhancement, and scaling to a size of 256×256 pixels. Image preprocessing was performed with OpenCV 4.1.2. This dataset was then randomly separated into two groups, 80% of the images were used for model training, and the remaining 20% for final model validation.

The training of the three models was performed in the same way. In all three cases, we started with a set of random weights adjusted with the SGD (Stochastic Gradient Descent) function as optimizer, and Categorical Cross Entropy as loss function. The training was performed over 80 epochs, and in each epoch, the accuracy and MSE (Mean Squared Error) values for both the training data and the validation data were checked. This process was used to make fine adjustments to the training, and to improve its performance in each case. All code was developed in Python 3.7 with support for Keras 2.6.0 and Tensorflow 2.6.0.

4. Results
The first results were obtained during the training process of the three models (Figures 3, 4 and 5). In all three cases, the models reached their highest performance point at a maximum of 40 epochs. Although in the following epochs a slight improvement was observed in all three cases, the benefit achieved does not justify the computational cost. The three models reached high Accuracy values for the training data (above 90%), however, they had different behaviors for the validation data. While the validation data were unknown for all three models, this tracking throughout training allowed the behavior to be adjusted to benefit the generalizability of the models. This worked very well with the ResNet and DenseNet models, but not with the NASNet model. The validation data for the ResNet and DenseNet models reached 30%, a value that, although low, exceeded the 10% achieved by the NASNet model. According to these behaviors, overfitting of the NASNet model is observed.

We calculated three metrics on the models using the validation data: Precision, Recall, and F1-score. The first one determines how well the model classifies items in a category concerning...
the total, the second one is more specific with the classified elements that actually belong to the
category, and the third one is a weighted average of the first two metrics. Tables 1, 2, and 3
show the results of these metrics. As observed from the accuracy, the ResNet and DenseNet
models perform much better with new data than the NASNet model. In fact, the NASNet model
classified all images in category 2, so it got right those that actually belonged to this category, but
missed all others. Consequently, the NASNet model does not meet the minimum performance
requirements. The ResNet and DenseNet models perform much better with validation data and
are very similar in each of the categories. This result validates them as promising in system
development and raises the question of the quality of the initial database.

Although the ResNet and DenseNet models achieve some generalization capability in
the presence of unknown data, their performance remains low compared to the expected
performance. Throughout the performance, the three models manage to reduce their error
and reach accuracy above 90% in all cases, but it is difficult for them to go above 30% with the
validation data. According to these results, it can be inferred that there is no overfitting of the
ResNet and DenseNet models, but that it is necessary to increase the complexity of the training
database to provide the models with more information when working with unknown datasets.
Table 1. Metrics by category and average. ResNet model.

|   | Precision | Recall | F1-score | Support |
|---|-----------|--------|----------|---------|
| 0 | 0.18      | 0.13   | 0.15     | 20      |
| 1 | 0.28      | 0.21   | 0.24     | 20      |
| 2 | 0.36      | 0.39   | 0.37     | 20      |
| 3 | 0.26      | 0.32   | 0.29     | 20      |
| 4 | 0.26      | 0.29   | 0.28     | 20      |
| 5 | 0.22      | 0.32   | 0.26     | 20      |
| 6 | 0.17      | 0.15   | 0.16     | 20      |
| 7 | 0.40      | 0.43   | 0.41     | 20      |
| 8 | 0.21      | 0.18   | 0.19     | 20      |
| 9 | 0.12      | 0.20   | 0.15     | 20      |
| 10| 0.38      | 0.59   | 0.47     | 20      |
| 11| 0.26      | 0.35   | 0.30     | 20      |
| 12| 0.44      | 0.30   | 0.36     | 20      |
| 13| 0.78      | 0.52   | 0.62     | 20      |
| 14| 0.44      | 0.22   | 0.29     | 20      |
| 15| 0.60      | 0.55   | 0.57     | 20      |
| 16| 0.27      | 0.21   | 0.24     | 20      |
| 17| 0.50      | 0.50   | 0.50     | 20      |
| 18| 0.21      | 0.25   | 0.23     | 20      |
| 19| 0.16      | 0.19   | 0.18     | 20      |
|   | Weighted avg. | 0.33 | 0.31 | 400 |

The last metric used on these models was the confusion matrix. Again, the validation data was used to evaluate the behavior of the models. To facilitate their analysis, a heat code was incorporated in which light colors indicate a higher quantity of elements, and dark colors a lower quantity; consequently, a well-defined light-colored diagonal should be observed in the best models. The results are shown in Figures 6, 7, and 8. As observed with the previous metrics, the best performances were achieved with the ResNet and DenseNet models, with DenseNet significantly better than ResNet, while it was verified that the NASNet model misclassifies most of the images in category 2. In general terms, the DenseNet model has an acceptable performance with validation data, and reduced size compared to ResNet, which makes it suitable for the development of the embedded identification system. Even so, the low capacity of the models to correctly infer categories for unknown data raises the need to evaluate the structure of the database used, probably by considerably increasing its size to increase the training data.

5. Conclusions

This paper presents the development of an autonomous model for the classification of high contact surfaces as a strategy for the application of specific maintenance and cleaning strategies in hotels and accommodation spaces with a high turnover of people, to ensure hygiene and health safety against possible disease contagion. The model is based on training from images containing characteristic elements of these environments such as electrical switches or doorknobs. According to the required performance, and the ability of the model to operate in real-time on an embedded device, we proposed the evaluation of three convolutional architectures that have reported high performance in this type of task. The ResNet, DenseNet, and NASNet topologies were selected and trained with their own dataset of 2000 images under the same conditions. The
Table 2. Metrics by category and average. DenseNet model.

| Precision | Recall | F1-score | Support |
|-----------|--------|----------|---------|
| 0         | 0.09   | 0.09     | 20      |
| 1         | 0.36   | 0.33     | 20      |
| 2         | 0.41   | 0.30     | 20      |
| 3         | 0.19   | 0.20     | 20      |
| 4         | 0.32   | 0.26     | 20      |
| 5         | 0.20   | 0.26     | 20      |
| 6         | 0.29   | 0.19     | 20      |
| 7         | 0.50   | 0.36     | 20      |
| 8         | 0.20   | 0.18     | 20      |
| 9         | 0.18   | 0.40     | 20      |
| 10        | 0.27   | 0.53     | 20      |
| 11        | 0.48   | 0.42     | 20      |
| 12        | 0.36   | 0.15     | 20      |
| 13        | 0.67   | 0.74     | 20      |
| 14        | 0.46   | 0.38     | 20      |
| 15        | 0.68   | 0.68     | 20      |
| 16        | 0.24   | 0.37     | 20      |
| 17        | 0.45   | 0.68     | 20      |
| 18        | 0.40   | 0.40     | 20      |
| 19        | 0.35   | 0.27     | 20      |
| Weighted avg. | 0.36 | 0.35 | 400 |

Figure 6. Confusion matrix. ResNet model.
Figure 7. Confusion matrix. DenseNet model.

three models were evaluated with the same metrics using a small set of the dataset not used in the training. The results show that the NASNet model is unable to generalize its behavior to unknown data, an activity better performed by the ResNet and DenseNet models. According to the results, the DenseNet model is selected as the architecture due to its high performance and smaller size compared to the ResNet model. Additional results include the need to optimize the DenseNet model in an embedded system with a larger and more complex database.
Table 3. Metrics by category and average. NASNet model.

|   | Precision | Recall | F1-score | Support |
|---|-----------|--------|----------|---------|
| 0 | 0.00      | 0.00   | 0.00     | 20      |
| 1 | 0.00      | 0.00   | 0.00     | 20      |
| 2 | 0.05      | 1.00   | 0.09     | 20      |
| 3 | 0.00      | 0.00   | 0.00     | 20      |
| 4 | 0.00      | 0.00   | 0.00     | 20      |
| 5 | 0.00      | 0.00   | 0.00     | 20      |
| 6 | 0.00      | 0.00   | 0.00     | 20      |
| 7 | 0.00      | 0.00   | 0.00     | 20      |
| 8 | 0.00      | 0.00   | 0.00     | 20      |
| 9 | 0.00      | 0.00   | 0.00     | 20      |
| 10| 0.00      | 0.00   | 0.00     | 20      |
| 11| 0.00      | 0.00   | 0.00     | 20      |
| 12| 0.00      | 0.00   | 0.00     | 20      |
| 13| 0.00      | 0.00   | 0.00     | 20      |
| 14| 0.00      | 0.00   | 0.00     | 20      |
| 15| 0.00      | 0.00   | 0.00     | 20      |
| 16| 0.00      | 0.00   | 0.00     | 20      |
| 17| 0.00      | 0.00   | 0.00     | 20      |
| 18| 0.00      | 0.00   | 0.00     | 20      |
| 19| 0.00      | 0.00   | 0.00     | 20      |
|   | Weighted avg. | 0.00 | 0.05 | 0.00 | 400 |

Figure 8. Confusion matrix. NASNet model.

Acknowledgments
The authors wish to acknowledge and thank the Technological Faculty of the Universidad Distrital Francisco José de Caldas for the support and funding during the course of this research work.

References
[1] Mendes M, Oliveira A, Pires O, Branca F, Beirao M, Santa A, Carvalho A and Alves J 2021 Acta Med Port 34 1–7 ISSN 0870-399X
[2] Brass A, Shoubridge A, Crotty M, Morawska L, Bell S, Qiao M, Woodman R, Whitehead C, Inacio M, Miller
