Recognizing Object Surface Materials to Adapt Robotic Disinfection in Infrastructure Facilities

Da Hu[1] **| Shuai Li**[1]

[1] Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, TN, 37996, USA

Correspondence
Shuai Li, Department of Civil and Environmental Engineering, The University of Tennessee, Knoxville, 851 Neyland Dr., Knoxville, TN, USA.
Email: sli48@utk.edu

Funding information
National Science Foundation, Grant/Award Numbers: 2026719, 1952140, and 2038967

ABSTRACT

Existing disinfection robots are not intelligent enough to adapt their actions to object surface materials for precise and effective disinfection. To address this problem, a new framework is developed to enable the robot to recognize various object surface materials and to adapt its disinfection methods to be compatible with recognized object surface materials. Specifically, a new deep learning network is proposed that integrates multi-level and multi-scale features to classify the materials on contaminated surfaces requiring disinfection. The infection risk of contaminated surfaces is computed to choose the appropriate disinfection modes and parameters. The developed material recognition method demonstrates state-of-the-art performance, achieving an accuracy of 92.24% and 91.84% on the Materials in Context Database (MINC) validation and test datasets, respectively. The proposed method was also tested and evaluated in the context of healthcare facilities, where the material classification achieved an accuracy of 89.09%, and the adaptive robotic disinfection was successfully implemented.

1 INTRODUCTION

Hospitals, nursing homes, airports, and buildings are hotbeds for pathogen colonization and transmission, resulting in a massive number of infections among the people who occupy these facilities (Lewis, 2021). Outbreaks of infectious diseases lead to illness and death, imposing significant burdens on the healthcare systems, reducing productivity, and leading to enormous economic losses. For example, the COVID-19 pandemic has led to over 221 million confirmed cases and 4.4 million deaths (Dong et al., 2020). The number of infections and deaths continues to increase with the emergence of more infectious variants of COVID-19, increasing the fear of future surging waves of infections. Healthcare facilities are particularly of concern during the pandemic given the influx of infected patients needing treatment. In healthcare facilities, surfaces can be contaminated through hand touching, respiratory droplets, or bodily secretions. This contamination can cause cross-transmission among patients and between patients and healthcare providers, jeopardizing people's health, and the normal operations of hospitals (Leung, 2021). In fact, before the COVID-19 pandemic, the United States Centers for Disease Control and Prevention (CDC) estimated that nearly 1.7 million patients are infected during hospitalization, resulting in 98,000 associated deaths (Haque et al., 2018). This statistic highlights the urgency and importance of proper surface disinfection to mitigate the transmission of infectious bacteria and viruses and to reduce the possibility and numbers of healthcare-acquired infections (HAIs) (Donskey, 2013).

Many facilities still rely on physical labor to carry out disinfection processes, such as using hydrogen peroxide and ultraviolet disinfection, which is time-consuming, labor-intensive, and poses an infection risk to the cleaning staff (Choi et al., 2021). Furthermore, manual disinfection is influenced by human behavioral factors, and real-world practices are highly variable (Doll et al., 2018). For instance, Rutala and Weber (2016) found that 10-50% of surfaces are contaminated in the rooms of patients infected with C. difficile, MRSA, and VRE. However, 51% of surfaces in patient rooms are found to not be thoroughly cleaned or disinfected, which could lead to a 120% increase of infection probability for future occupants of the room. There is a critical need for intelligent robotic disinfection to reduce viral bioburdens on contaminated surfaces, and thus prevent fomite-mediated transmission of infectious pathogens.

The elevated concerns due to the COVID-19 pandemic have increased the adoption of robotic technology for infection control and environmental hygiene (Zemmar et al., 2020). Existing robotic disinfection technologies are subject to two main limitations. First, existing disinfection robots are perceived as roaming bases with disinfection sources such as UV lights for coarse disinfection. This requires the absence of people in the rooms or buildings to be disinfected (Diab-El Schahawi et al., 2021). Given the high volume of patients in healthcare facilities needing treatments, these disinfection robots are difficult to deploy. Second, healthcare facilities harbor a variety of pathogens that colonize a wide spectrum of surfaces made from different materials. The object surface materials have significant impacts on pathogen colonization and transmission, and thus require different disinfection modes, parameters, and procedures to ensure complete and efficient disinfection. For example, a recent study by Chin et al. (2020) suggested that certain pathogens, such as SARS-CoV-2, can stay infectious for as long as 7 days on metal and plastic surfaces, while SARS-CoV-2 may survive for only 2 days on fabric (2020). Furthermore, the transfer efficiency or transmission rate of bacteria or viruses to hands from a surface differs between materials (Lopez et al., 2013). For example, the transfer efficiency of MS2 can reach 19.3% on glass surfaces but only 0.3% on fabric surfaces under a relative humidity of 15% to 32% (Lopez, et al., 2013). Therefore, the materials of a contaminated object surfaces must be considered for appropriate robotic disinfection to occur. However, this research topic has not yet been investigated.

To address the challenges, a disinfection robot prototype was proposed in our prior study (Hu et al., 2020a) and is under continuous development, as presented in Figure 1. This robot, with a mobile base, a six-degree-of-freedom robotic arm, and sensors including LiDAR and RGB-D cameras, can navigate in buildings, recognize and segment potentially contaminated areas, and create 3D semantic maps for precise and intelligent disinfection. A UV light disinfection source is mounted on the robotic arm end-effector with a reflective shield. The end-effector can be further customized with other disinfection modes such as “spray” and “wipe”. To enable the adaptive disinfection with multiple disinfection modes and to optimize the disinfection parameters with respect to the object surface materials that mediate the colonization and transmission of varying pathogens, the robot needs to be endowed with a robust capability to rapidly and accurately recognize the materials of the contaminated object surface requiring disinfection, and to computationally link the robotic perceptions, i.e. material recognition, with robotic actions, i.e. disinfection modes, procedures, and parameters.

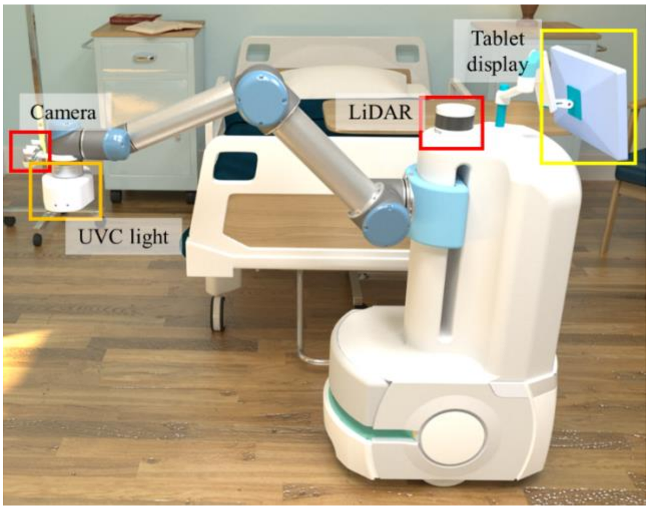


FIGURE 1. Illustration of the proposed disinfection robot prototype

The objective of this study is to develop a new computational process and deep-learning-based method to enable the robot to recognize the material types of object surfaces and to adapt disinfection modes and parameters to achieve a complete and efficient disinfection. This study features two contributions to the body of knowledge, with the potential to transform the current paradigms of robotic disinfection in infrastructure facilities.

- First, a new deep-learning network is proposed to classify the materials of object surfaces needing disinfection. Our designed network innovatively integrates multi-level Convolutional Neural Network (CNN) features, multi-scale CNN features, and a texture encoder network in an end-to-end learning fashion, which has not been integrated by existing studies. The multi-level CNN features can capture high-level abstract representations of the material and low-level texture and color information, which can enhance material representation ability of the network. The multi-scale features are captured by the Atrous Spatial Pyramid Pooling (ASPP) with multiple resampling rates, allowing the network to learn spatial repetitive features of material textures. The texture encoder network can capture texture details and local spatial information from different levels. The orderless features and ordered spatial information are then balanced with a bilinear model. The proposed network extracts rich features for accurate material representation, achieving state-of-the-art results on six public material datasets.
- Second, facilities such as hospitals are highly heterogeneous with different types of units serving distinct functionalities and characteristics, providing habitats for various bacteria and viruses with different

transmission and infection pathways. A new method was proposed in this study to computationally link the recognition of materials with the robotic disinfection actions, completing the loop from robotic perception to robotic actions. Our innovation lies in the computational modeling of the interactions among surface materials, pathogens, and disinfection modes and parameters to adapt robotic disinfection actions, which has not been achieved by existing studies and current systems. The developed methods could lead to an intelligent robotic disinfection paradigm that goes well beyond existing systems that are perceived as roaming UV lights for coarse disinfection. Intelligent and precise robotic disinfection can be implemented in critical infrastructure facilities such as hospitals, airports, school buildings, and food processing plants to improve environmental and public health.

2 BACKGROUND AND LITERATURE REVIEW

In this section, related studies on robotic disinfection and material recognition are reviewed, and the existing knowledge gaps are summarized.

2.1 Related studies on robotic disinfection

Robotic disinfection has long been treated as a solution to mitigating the spread of infectious diseases in infrastructure facilities, which has been an active research area in recent years. The disinfection mode of these disinfection robots can be characterized by “UVC light”, “wipe,” and “spray”.

A UV-disinfection robot offers a non-touch method, disinfecting surfaces from a distance using a UVC light. UVC light is an environmentally friendly disinfection method, as it does not leave any residues on surfaces. UV-disinfection robots are essentially mobile robots with UVC light columns mounted to their top. These robots are commonly integrated with a variety of sensors for navigation and object detection, such as cameras, LiDAR, and ultrasound. Gibson et al. (2017) deployed Xenon UV-disinfection robots in a hematopoietic stem cell transplant unit for three months and found the rate of HAI decreased to less than one per quarter. In (Fleming et al., 2018), UV-disinfection robots were deployed facility-wide for terminal disinfection of the rooms that hospitalized patients infected with Clostridium difficile. Since the beginning of the COVID-19 pandemic, many UV-disinfection robots have been developed and tested. For example, the MIT UV robot consists of an Ava robotics’ mobile base and a customized UVC light fixture, and the robot’s disinfection capabilities were tested in a food bank (Gordon, 2020). The robot took around 30 minutes to cover 4,000-square-foot spaces with a speed of 0.22 mph. The power and the number of light columns could be customized based on the size of rooms and their disinfection requirements. However, since UVC lamps are powerful enough to cause harm to the skin and eyes, the rooms must be evacuated during disinfection. To overcome this limitation, McGinn et al. (2021) developed a prototype UV-disinfection robot called the “Violet robot platform”. A UVC reflectance shield was added to enclose the UVC lamp, thus reflecting the radiation emitted behind the robot. The authors claim that the Violet robot has the potential to work safely alongside human cleaners. In Hu, et al. (2020a), a UVC light wand was mounted onto a robotic arm with a mobile base to navigate in the built environment and to disinfect potentially contaminated surfaces. The major drawback of the existing UV-disinfection robots is that they cannot perceive object surface materials and adapt their disinfection parameters, which may lead to the incomplete disinfection of high-risk surfaces.

Wiping with chemical disinfectants is the main disinfection method for decontamination of high-touch surfaces in infrastructure facilities. “Wipe” mode has also been integrated into robot systems to disinfect contaminated surfaces. Toyota developed a ceiling-mounted home robot to wipe surfaces with soft rubble mounted on the gripper (Vincent, 2020). The robot can travel on the ceiling to avoid the problems associated with navigating a cluttered floor. Ramalingam et al. (2020) proposed a disinfection robot prototype to automate the disinfection of doorhandles in infrastructure facilities. The authors designed a deep learning model to detect doorhandles in the image and to calculate the doorhandle location. More recently, the Fraunhofer Institute for Manufacturing Engineering and Automation proposed a prototype robot, called DeKonBot, to disinfect contaminated surfaces such as bedrails, light switches, and elevator buttons (NOVUS, 2021). The DeKonBot consists of a mobile base and a robotic arm that carries out a wipe disinfection mode. Material information is critical for the wipe-disinfection robot because wipe mode is not suitable for certain types of surfaces, such as fabric and paper. However, existing wipe-disinfection robots have not been developed with the capability to recognize the material of the surfaces that require disinfection.

The disinfectant spray is another important disinfection method that has been widely used in the COVID-19 pandemic. For example, a smart prefabricated sanitizing chamber was designed for COVID-19 to enable a uniform spraying of sanitizing fluid onto healthcare workers (Abu-Zidan et al., 2021). The spray mode has also been integrated with the robotic system. Zhao et al. (2021) developed a smart disinfection robot system that sprays disinfectants in the operating theaters or the patients’ rooms in healthcare facilities. The developed system primarily focused on the integration of multiple technologies, such as the Internet of Things (IoT), SLAM, hand gesture recognition, and navigation. Thakar et al. (2021) developed an area-coverage planning algorithm for a spray-based disinfection robot to compute a path for the nozzle to follow to completely disinfect surfaces.

The remote operator needs to select the area to be disinfected and extract the corresponding point cloud for path calculation. These spray-based robots achieved good results in controlled experiments. However, the disinfection dose is not adaptable based on the surface materials, which can lead to incomplete disinfection. Furthermore, spray-based disinfection could damage paper surfaces. Therefore, it is important for these spray-based robots to recognize surface materials and to adapt disinfection modes and parameters.

2.2 Related Studies on material recognition

Material and texture recognition is challenging, and numerous algorithms have been developed to address this challenge. Traditional material recognition methods relied on handcrafted features that are usually not robust and are computationally intensive due to the high dimensions (Caputo et al., 2005). To address these limitations, CNN-based methods were applied to learn material features in images. Compared to handcrafted features, automatically learned features are found to be more robust in image classification (Kang et al., 2018). In (Cimpoi et al., 2016), the CNN-based feature extraction was demonstrated to be efficient for material and texture recognition. In their method, CNN was truncated at the level of the convolutional layer to obtain so-called local image descriptors. Traditional orderless pooling encoders (e.g., Fisher Vector (FV); Vector of Locally Aggregated Descriptors (VLAD); BOW) were used to encode CNN features to a feature vector for classification. The feature vector was then passed to an SVM model to predict the material labels. A combination of FV and CNN (FV-CNN) achieved superior results in image classification. In a follow-up study by Song et al. (2017), the output from FV-CNN was further refined through learnable locally connected layers. There are two major drawbacks of this method. First, the method needs a large disk space to store the features extracted from CNN. Second, each step is trained separately, which means the training of the material classifiers cannot update CNN model weights, comprising both training efficiency and accuracy.

With the advancement of computational power, many advanced CNN methods have been developed for applications in different research fields, such as vehicle detection (Arabi et al., 2020), human activity recognition (Luo et al., 2020), pavement condition assessment (Hsieh et al., 2021), and earthquake early warning systems (Rafiei and Adeli, 2017). In the domain of material and texture recognition, a variety of end-to-end CNN models have been developed. For instance, Bell et al. (2015) investigated the performance of three popular CNN architectures for material recognition: AlexNet, VGG-16, and GoogLeNet. The fine-tuned AlexNet yielded good results on the Flickr Material Database (FMD). In Xue et al. (Xue et al., 2017), a CNN-based Differential Angular Imaging Network (DAIN) was developed to integrate multi-view images to recognize outdoor materials. Original and differential angular images were fed into the network and their final prediction results were combined. The results indicated that the prediction accuracy was significantly increased with differential angular images. However, typical CNNs with fully connected (FC) layers cannot capture spatially invariant features on materials. To address this issue, Zhang et al. (2017) developed the Deep Texture Encoding Network (Deep-TEN) with an orderless feature pooling encoder network. The encoding layer is integrated into the CNN network to learn visual dictionary extracts from CNN features. Building upon the Deep-TEN, Deep Encoding Pooling Network (DEP) was designed to integrate high-level spatial information and orderless features for the task of material recognition (Xue et al., 2018). The DEP added a pooling layer in combination with the texture encoder. A bilinear model was then used to merge the outputs from the pooling layer and texture encoding layer.

In recent years, Zhai et al. (2019) proposed a Deep Multiple-Attribute-Perceived Network (MAPNet) to perceive multiple visual attributes for texture recognition. The MAPNet was based on a multi-branch architecture that allows for visual texture attributes to be learned synergistically. The CNN features from each branch are fed into a spatially adaptive global average pooling for feature aggregation. In Zhai et al. (2020), a Deep Structure-Revealed Network (DSRNet) was developed by leveraging spatial dependency among the captured primitives as structural representations. The DSRNet devised a primitive capturing module to generate primitives from different directional spatial contexts. The primitives were fed into a dependence learning module to learn structural representations for material recognition. Chen et al. (2021) aggregated cross-layer statistical self-similarity information as a feature aggregation module in their CLASSNet network. Feature maps across different layers are modeled as a dynamic process with self-similarity statistics. The cross-layer statistics are characterized using a histogram of cross-layer features based on differential box-counting. This method achieved state-of-the-art performance for the time on several public material datasets. The review of material recognition methods demonstrates that CNN methods are robust and powerful for material representation.

2.3 Knowledge gaps

Two main knowledge gaps are identified in the related studies on robotic disinfection and material recognition. First, object surface materials play a significant role in pathogen persistence and transmission. It thus follows that there is a need for material-specific disinfection parameters to ensure complete and efficient disinfection. However, the existing robotic disinfection systems mentioned above have not yet considered the object surface materials in guiding their disinfection tasks. This could lead to two undesirable consequences: (1) the robot may not completely disinfect high-risk surfaces; (2) the disinfection mode may not be suitable for certain object surfaces (e.g., a wipe-based robot is not suitable for fabric and paper surfaces). There still lacks a computational model that considers the interactions of human-pathogen-fomite and links surface materials with robotic disinfection modes and parameters.

Second, recognizing materials is challenging due to a diverse range of appearances and spatially invariant features. Existing studies, such as Deep-Ten (Zhang, et al., 2017) and (Xue, et al., 2018), aim to capture spatially invariant features of materials with the integration of an orderless feature pooling layer in an end-to-end learning fashion. The DSRNet (Zhai, et al., 2020) is another network that periodically captures recurrent features by learning the inherent spatial dependency of multiple primitives from different directions. However, their networks cannot capture low-level texture and color information, which is important for material recognition tasks. In the latest study, CLASSNet (Chen, et al., 2021) utilized the information from different layers which achieved state-of-the-art performance on material recognition, even though spatially invariant features were not captured. No study to our knowledge has utilized the multi-scale features for material representations. Furthermore, there lacks a network to learn both multi-level and multi-scale features simultaneously and to encode these features in an orderless manner to capture spatially invariant features.

3 METHODOLOGY

The disinfection robot can navigate in a building and recognize potentially contaminated areas based on our developed method (Hu, et al., 2020a). The robot then moves to the proximity of the contaminated objects needing disinfection and adapts UVC light scanning trajectories. The limitation of the robot is the lack of capability in recognizing surface materials and computationally linking these materials to robotic disinfection parameters, which impacts its disinfection efficiency. This study aims to address this limitation by developing a new computational process and material recognition network to enable the robot to adapt its disinfection modes and parameters. Figure 2 presents an overview of the research framework of this paper. First, a deep learning network is proposed to recognize the object surface materials captured by the disinfection robot. The proposed network integrates multi-level CNN features, which leverages both low and high-level information to capture semantic and texture information. High-level features can capture semantic features, which are abstract representations of the material. Low-level features can capture more subtle details such as texture information. The Atrous Spatial Pyramid Pooling (ASPP) module can extract multiscale features by resampling feature maps at multiple rates. The ASPP module increases the size of the receptive field without compromising the feature map resolution. Our network further integrates an encoder component, which combines both orderless and local spatial feature pooling. The encoder can preserve texture and ordered spatial information from different layers, which can better capture spatially invariant features of materials.

Second, a fomite transmission model considers fomite-pathogen-human interactions to compute the risk of infection from object surfaces given the recognized material. The estimated infection risk is then used to determine the $\log_{10}$ reduction needed to reduce viral bioburden on the object surface to below safety target levels. The required $\log_{10}$ reduction in bioburden is used to optimize the disinfection parameters (i.e., exposure time, distance, and irradiance) for the robot to implement. Finally, experiments are conducted in three parts. First, disinfection mode and parameters for different object surfaces are analyzed. Second, the proposed material recognition network is evaluated on the Material in Context Database (MINC) and an additional material dataset collected in the context of healthcare facilities. The network is then compared to other state-of-the-art methods followed by an ablation study to evaluate the effectiveness of each module in the network. Third, the adaptive robotic disinfection framework is implemented in a fully modeled patient room.

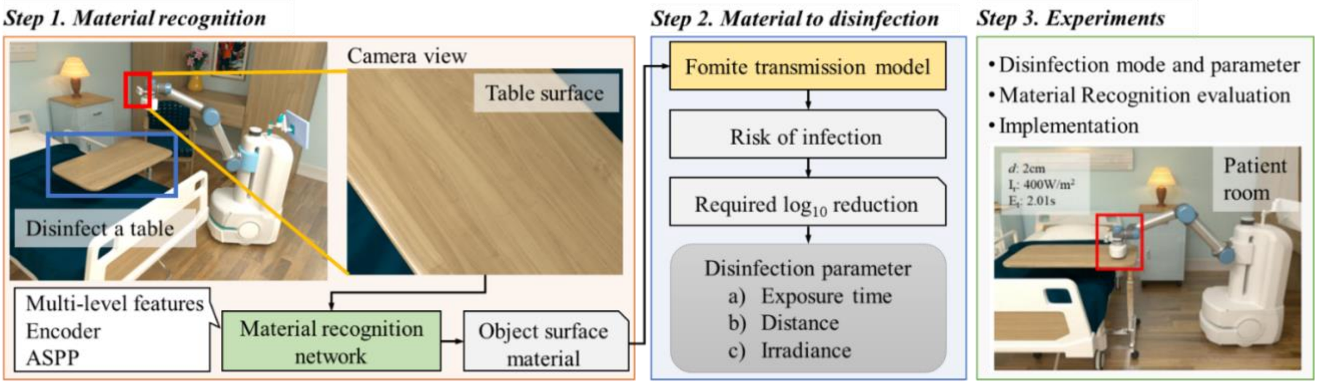


FIGURE 1. Overall research framework of this paper

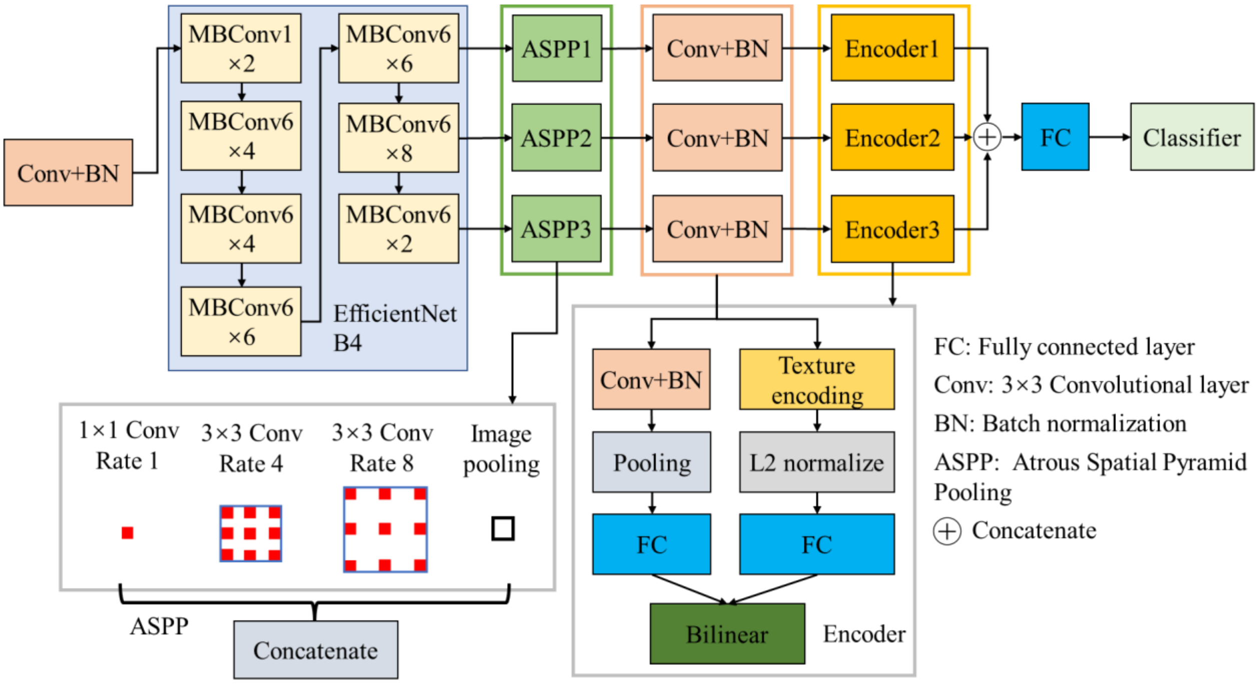


FIGURE 3. Flowchart of the proposed network

3.1 Material recognition network

Material recognition is needed to determine a suitable disinfection method for a surface. A deep learning-based classification network is proposed to recognize the materials of contaminated surfaces. Figure 3 presents an illustration of the proposed network, which is composed of four components: the backbone, a multi-level feature fusion, Atrous Spatial Pyramid Pooling (ASPP), and an encoder. Each component is detailed below.

Backbone. The classification model is designed based on the EfficientNet-B4 network proposed in Tan and Le (2019). The EfficientNet-B4 network is small and fast on inference. The input is first fed into a 3X3 convolution, batch normalization, and activation layer. The outputs are then fed into 7 inverted residual blocks, also known as MBConv blocks (Tan and Le, 2019), optimized by the squeeze-and-excitation method (Hu et al., 2020b). MBConv[N] represents an MBConv with an expansion factor of N. MBConv1 is a depth wise separation block without the expansion operation. The MBConv6 block is the inverted residual block with an expansion factor of 6. The number of sub-blocks for the 7 MBConv blocks are 2, 4, 4, 6, 6, 8, and 2, respectively.

Multi-feature integration. The multi-level features are innovatively extracted in this study to capture the low-level texture and color information and the high-level semantic information. Specifically, the outputs from the last three MBConv blocks are extracted and separately fed into the ASPP component. The multi-level CNN features are extracted to utilize features from different layers of the EfficientNet-B4 network. The reason for multi-layer feature fusion is that texture details learned from the shallow layers tend to vanish with going deeper into the layer . The texture details learned from the low-level features are important for material recognition. The features from multiple layers can capture complementary information and a combination of these features can improve performance.

ASPP. The ASPP is used to obtain multi-scale context information (Chen et al., 2018). The outputs from the last three MBConv blocks are separately fed into the three ASPP layers. The ASPP layer consists of three Atrous convolutions with rates of 1, 4, and 8 and one global average pooling layer. Different rates of Atrous convolution have different sizes for their receptive fields. Since material textures are typically translationally invariant, a larger size for the receptive field can better capture spatial repetition features. The ASPP layer can extract multi-scale features while preserving the resolution of the features. The features extracted from multiple rates and the pooling layer are fused as the global features. The kernel sizes for Atrous convolutions are 1×1, 3×3, and 3×3. Atrous convolution is a generalized standard convolution and expands the window size to capture large features without adding computational cost by inserting zero-values into the convolution kernels. The outputs of the ASPP layer are then fed into a 3×3 convolutional layer with batch normalization.

Encoder. Since material properties are usually translationally invariant, material recognition methods need to capture an orderless measure encompassing some spatial repetition. Previous studies have shown that orderless pooling, like the Fisher Vector (FV), works better than order-sensitive pooling in material recognition (Cimpoi, et al., 2016). The CNN combined with orderless pooling encoders, such as BOW, FV, and Vector of Locally Aggregated Descriptors (VLAD), has been demonstrated to be effective in material classification (Cimpoi, et al., 2016). The proposed encoder module consists of a texture encoding network (TEN) (Zhang, et al., 2017) and local spatial pooling (LSP).

The TEN is used to build dictionary learning to extract visual codewords. The TEN can encode CNN features in an orderless manner, such as FV and VLAD, using a residual layer. The input of the TEN component is the output from the ASPP module with a shape of C×H×W, where C is the dimension and H×W represents the size of the ASPP outputs. The feature map is formed as a C-dimensional feature vector X = {x_1, … x_m}, where m represents the feature number for each channel. The TEN layer learns an inherent codebook C = {c_1, … c_k} and smoothing factors S = {s_1, … s_k}. The definition of residual encoding vector for the codeword c_j is given in Eq. (1), where r_{ij} is the residual vector calculated as $r_{ij} = x_i - c_j$. The TEN layer aggregates CNN features into residual encoding vectors E = {e_1, … e_k}. Note that increasing the number of codewords has the potential to capture more detailed texture information.

$$e_j = \sum_{i=1}^{m} \frac{e^{-s_j\left\|r_{ij}\right\|^2}}{\sum_{n=1}^{k} e^{-s_n\left\|r_{in}\right\|^2}} r_{ij} \tag{1}$$

To capture spatial information, the outputs from the ASPP module are fed into a 3×3 convolutional layer with a stride of 2 and a batch normalization operation to standardize the outputs. A two-dimensional adaptive average pooling is then applied over the outputs from the batch normalization. A fully connected (FC) layer is used for feature dimension reduction. A bilinear model is used to fuse outputs from the TEN and LSP by multiplying their feature maps using the outer product. The outer product captures the correlations of orderless features and local spatial descriptors. The bilinear function is given by Eq. (2), where $\mathbf{u} \in \mathrm{R}^{d}$ is the output from the TEN network, $\mathbf{v} \in \mathrm{R}^{d}$ is the output from the LSP layer, $\boldsymbol{f} \in \mathrm{R}^{dxd}$ is the output of the bilinear model.

$$\boldsymbol{f} = \boldsymbol{u} \otimes \mathbf{v} \tag{2}$$

Table 1 displays the detailed architecture of the proposed network. The input image size is 224x224x3. The outputs dimension from the Efficientnet_B4 module is 448x7x7, which is fed into the ASPP modules. The dimension of the outputs from the ASPP1, ASPP2, and ASPP3 are 512x14x14, 1024x7x7, and 1024x7x7, respectively. The channels of the outputs from the ASPP modules are reduced to 384 using a convolutional layer. The outputs are then passed to the Encoder modules, which consist of the TEN and LSP layers. There are 8 codewords for the TEN layers. The feature dimensions from the TEN is 1x3072, which is then fed into an FC layer to reduce the feature dimension to 1x64. The LSP first applies a convolutional and batch normalization operation, and then the average pooling operation is conducted. An FC layer is applied as a dimension reduction step for outputs from the pooling layer. The dimensions of outputs from the TEN and the LSP are both 1x64. The bilinear model is then used to fuse the outputs from these two modules together, with an output of 1x4096. The outputs from Encoder1, Encoder2, and Encoder3 are concatenated together as a 1x12288 feature vector. Note that L2 normalization is used for the outputs from the TEN layer and the bilinear model. Finally, a fully connected classifier is used to classify the image.

TABLE 1. Architecture of the proposed network. 'KS' is kernel size; 'OS' is output size; '#CN' is the number of channel; '#LN' is the number of layer.

Module	Operator	KS	OS	#CN	#LN
EfficientNet-B4	Conv+BN+Swish	3x3	224x224	48	1
	MBConv1	3x3	112x112	24	2
	MBConv6	3x3	112x112	32	4
	MBConv6	5x5	56x56	56	4
	MBConv6	3x3	28x28	112	6
	MBConv6	5x5	14x14	160	6
	MBConv6	5x5	7x7	272	8
	MBConv6	3x3	7x7	448	2
ASPP1	Atrous Conv1	1x1	14x14	128	1
	Atrous Conv2	3x3	14x14	128	1
	Atrous Conv3	3x3	14x14	128	1
	Pooling+Conv+BN+Relu	1x1	14x14	128	1
	Concatenate	-	14x14	512	1
	Conv+BN	1x1	14x14	384	1
ASPP2 & ASPP3	Atrous Conv1	1x1	7x7	256	1
	Atrous Conv2	3x3	7x7	256	1
	Atrous Conv3	3x3	7x7	256	1
	Pooling+Conv+BN+Relu	1x1	7x7	256	1
	Concatenate	-	7x7	1024	1
	Conv+BN	1x1	7x7	384	1
Encoder1	TEN	-	1	3072	1
	FC	-	1	64	1
	Conv+BN	3x3	6x6	192	1
	Pooling	6x6	1x1	192	1
	FC	-	1	64	1
	Bilinear mapping	-	1	4096	1
Encoder2 & Encoder3	TEN	-	1	3072	1
	FC	-	1	64	1
	Conv+BN	3x3	3x3	192	1
	Pooling	3x3	1x1	192	1
	FC	-	1	64	1
	Bilinear mapping	-	1	4096	1
	Concatenate	-	1	12288	1
Classifier	FC	-	1	256	1
	Classification	-	1	9	1

3.2 From material to disinfection

In this section, we elaborate on how to enable the robot to adapt its disinfection mode and parameters by utilizing the material information. The fomite transmission risk for different surfaces is critical in determining the disinfection practice needed to prevent the spread of pathogens. For surfaces that have a higher infectivity risk, a higher disinfection level is needed for complete disinfection.

The concentration of bacteria and viruses undergoes an exponential decay in the survival period on different surfaces (Guo et al., 2021). The concentration of bacteria and viruses at different times is given in Eq. (3), where C_0 is the initial concentration, t is the time in hours, and λ is the inactivation rate.

$$C_h = C_0 \times e^{-\lambda t} \tag{3}$$

The parameter λ can be estimated based on pathogen survivability on surfaces. In Chin, et al. (2020), SARS-CoV-2 was found to not be detectable when the concentration was smaller than 100 TCID50/mL, which is around 0.02% of the initial concentration on wood surfaces. Therefore, 0.02% is used as the survival fraction for pathogens at the end of the survival periods shown in Table 2. The parameter λ can be calculated as ln5000/st, where st is the survival time of the pathogen.

A stochastic-mechanistic model developed by Pitol and Julian (2021) is adapted in this study to estimate the infection risk from contaminated surfaces with different materials. The model is built based on surface-to-hand-to-mucous contact. The transfer of a pathogen from surface-to-hand is first calculated in Eq. (4), where C_h is the bacteria and virus concentration on the surface at time h, c_*PFU* is the conversion factor from Genome copies and Colony-forming unit to the infectious virus in PFU, *eff* is the pathogen recovery efficiency from surfaces, TE_{sh} is the transfer efficiency of the pathogen from surface to hand, α is $\log_{10}$ reduction in the number of bacteria and virus, and C_{hand} is the concentration on the hand.

$$C_{hand} = \frac{C_h}{eff} \times c\_PFU \times TE_{sh} \times \frac{1}{10^{\alpha}} \tag{4}$$

The transfer of the pathogen from hand-to-mucous can be approximated by the concentration of pathogens on the hand and the transfer efficiency, which is defined in Eq. (5), where TE_{hm} is the transfer efficiency from the hand to mucous, and *FSA* represents the fractional surface area in contact with the mucous membranes, and D is the infectious dose.

$$D = C_{hand} \times TE_{hm} \times FSA \tag{5}$$

The infectious dose is then used to estimate the risk of infection using Eq. (6), where k represents the dose-response parameter.

$$P = 1 - e^{-kD} \tag{6}$$

The $\log_{10}$ reduction of the viral bioburden on surfaces represents the level of disinfection needed, with a greater value representing a higher disinfection level. The risk threshold is set to 10^{-6}, which is the U.S. EPA threshold for water quality, based on various representative infectious diseases (Boehm, 2019). The robot needs to adapt disinfection parameters to lower the infection risk below 10^{-6}. The required $\log_{10}$ reduction is an input parameter to calculate the disinfection parameters for our UV-disinfection robot (Hu, et al., 2020a) with the applicable exposure time, distance, and irradiance. The bacteria and virus decay when exposed to UVC light can be estimated as a first-order decay rate model (Kowalski, 2009), which is defined in Eq. (7), where α is $\log_{10}$ reduction needed to lower infection risk below 10^{-6}. ρ is the UVC inactivation rate (m^2/J), E_D is the UVC exposure dose (J/m^2).

$$\frac{1}{10^{\alpha}} = e^{-\rho \times E_D} \tag{7}$$

The inactivate rate ρ is 0.0343 m^2/J for coronavirus (Bouri and Shatalov, 2020). The UVC exposure dose E_D is calculated in Eq. (8), where I_r (W/m^2) is the UVC irradiance at 1cm, E_t is the UVC exposure time in seconds, and d represents the distance between UVC light and object surfaces.

$$E_D = \frac{1}{d^2} \times I_r \times E_t \tag{8}$$

4 EXPERIMENT AND RESULTS

This section details the analysis of disinfection mode and parameters, evaluation on material recognition network, and implementation.

4.1 Analysis of disinfection mode and parameters

4.1.1 Experiment settings

Healthcare facilities contain different types of units that serve patients battling with different illnesses, providing habitats and transmission pathways for various infectious pathogens. In this study, SARS-CoV-2 and Escherichia coli (E. coli) are selected as the representative examples to illustrate how to transfer material information to disinfection practices. SARS-CoV-2 continues to lead to outbreaks of COVID-19 in healthcare facilities. E. coli has been identified as the major cause of urinary tract infections in healthcare facilities (Bergeron et al., 2012). Patients infected with SARS-CoV-2 and E. coli are typically diagnosed in the pulmonology and urology departments and are hospitalized in different patient rooms. Therefore, the disinfection practices need to be adapted according to the prevalent infectious pathogens present in different types of hospital rooms.

Table 2 presents the survival time and transfer efficiency for SARS-CoV-2 and E. coli on different surfaces. As indicated, SARS-CoV-2 and E coli. can generally survive longer on hard surfaces than on soft surfaces. For instance, SARS-CoV-2 can stay active for 7 days on metal and plastic but only 3 hours on paper (Chin, et al., 2020). Note that the survival time of SARS-CoV-2 is used for each material except for ceramic. The survivability of HCOV-229E, which is also a species of coronavirus, on ceramic is used instead.

TABLE 2. The survival time and transfer efficiency for SARS-CoV-2 and E. coli

Surface type	Material	SARS-CoV-2		E. coli	
		Survival time	Transfer efficiency (%)	Survival time	Transfer efficiency (%)
Soft	Fabric	2 days [1]	0.73[2]	4 – 56 days [3]	5.32[2]
	Leather	1 day [4]	7.00[5]	-	-
	Paper	3 hours [1]	0.55[2]	1 – 96 hours [6]	0.08[2]
Hard	Ceramic	5 days [7]	24.15[2]	14 days [8]	36.15[2]
	Glass	4 days [1]	43.30[2]	1 – 14 days [3]	41.85[2]
	Metal	7 days [1]	21.95[2]	14 – 60 days [3]	28.95[2]
	Plastic	7 days [1]	50.60[2]	24 h – 300 days [3]	47.00[2]
	Polished stone	5 days [9]	20.10[2]	-	21.9[2]
	Wood	2 days [1]	31.50[10]	2h – 28 days [3]	-

Note: [1,4,9] are sources of survival time for SARS-CoV-2 (Chin, et al., 2020; Anderson and Boehm, 2021; Virtanen et al., 2021). [7] provides the survival time of HCOV-229E on ceramic (Warnes et al., 2015). [3] is sources of survival time for E. coli (Wißmann et al., 2021). [6] provides the survival time of Francisella tularensis on paper (Richter et al., 2019). [8] is the survival time of Klebsiella pneumoniae on ceramic (Esteves et al., 2016). [2] is the source of transfer efficiency for MS2 and E. coli (Lopez, et al., 2013). Refs. [5,10] are sources of transfer efficiency for MS2 (Anderson and Boehm, 2021; Castaño et al., 2021)

For fomite-to-hand transfer efficiency, data from MS2 coliphage are used due to the unavailability of SARS-CoV-2 transmission data. MS2 and SARS-CoV-2 are both single-stranded RNA viruses, which have similar transfer mechanisms from fomite to humans. Furthermore, MS2 has been used as a surrogate to facilitate the investigation of transmission and disinfection of SARS-CoV-2 (Castaño, et al., 2021). Table 2 also presents the survival time and transfer efficiency used for E. coli. Note that for the survival time of E. coli on paper and ceramic, Francisella tularensis and Klebsiella pneumoniae are used, as they are both gram-negative bacteria like E. coli.

Data collected at different surfaces indicated that the concentration of SARS-CoV-2 varied from 0.1 to 102.4 gc/cm^2 (Harvey et al., 2020; Abrahão et al., 2021). For E. coli, the concentration varied from 0.1 to 15.8 CFU/cm^2 (Trindade et al., 2014; Cinar and OnbaŞI, 2021) on contaminated surfaces. In this study, the initial concentration of SARS-CoV-2 and E. coli are assumed to be 100 gc/cm^2 and 10 CFU/cm^2, respectively. Table 3 gives the input parameters and their distributions used to estimate required $\log_{10}$ reduction.

TABLE 3. Input parameters

Parameter	Unit	SARS-CoV-2	E. coli
TE_{hm}	unitless	Normal (0.20, 0.06) [a]	
FSA	cm^2	Uniform (4, 6) [b, c]	
k	PFU^{-1}	Triangle (0.00107, 0.00246, 0.00680) [d]	
eff	unitless	Normal (0.6, 0.266)[e]	
c_PFU	unitless	Uniform (0.01,0.001) [d]	Uniform (0.01,0.05) [f]

Note: [a] is ref. (Pitol et al., 2017); [b] is ref. (AuYeung et al., 2008); [c] is ref. (Pitol and Julian, 2021); [d] is ref. (Kraay et al., 2018); [e] is ref. (Harvey, et al., 2020); [f] is ref. (Mudgal et al., 2006).

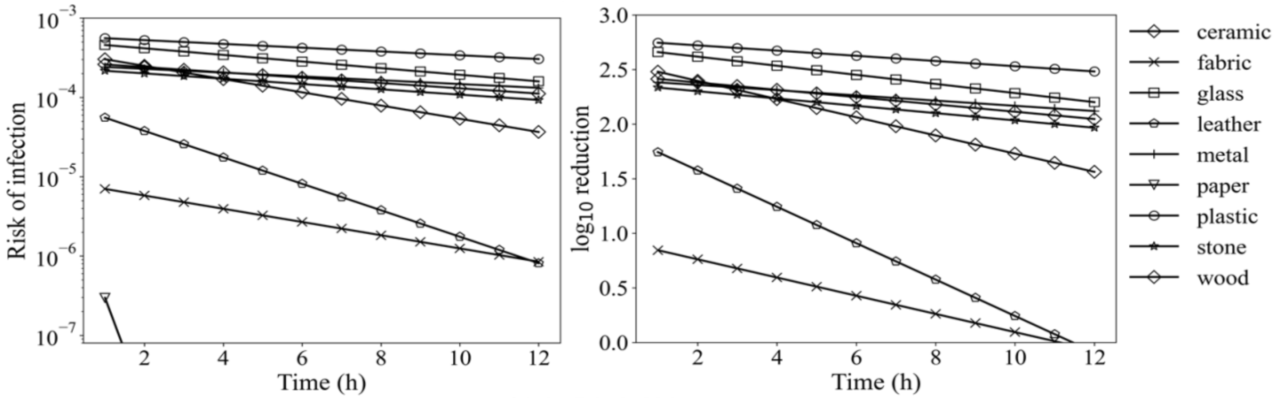
4.1.2 Analysis of results

The Monte Carlo simulation is used to estimate the infection risk by incorporating the input parameters' distributions. The model is simulated 50,000 times and the median risk values are reported. The survival time of E. coli is assumed to be a uniform distribution within the range. Figure 4 shows the estimated infection risk for different surfaces and the required $\log_{10}$ reduction to lower risk to below 10^{-6}. The results indicate that infection risks of SARS-CoV-2 and E. coli for plastic, glass, metal, ceramic, and polished stone are higher than 10^{-4} within 12 hours. Soft surfaces, such as leather and plastic, have a lower infection risk compared to hard surfaces, such as plastic, glass, and metal. In addition, paper surfaces show a low infection risk, which is smaller than 10^{-6}.

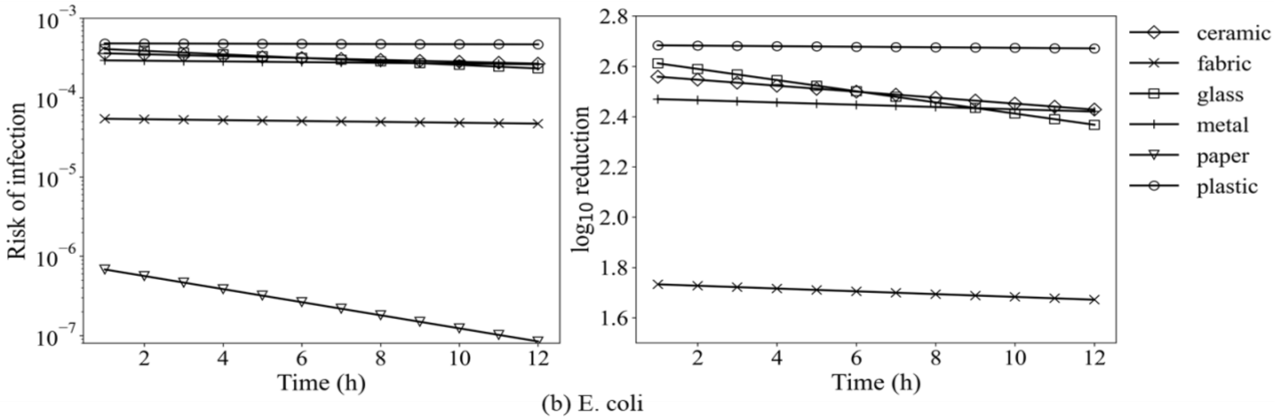
In cleaning and disinfection practices, the disinfection dose should be higher than needed to meet the disinfection requirements (Collivignarelli et al., 2018). Therefore, the $\log_{10}$ reduction is rounded up to an integer to ensure the object surfaces are completely disinfected. As indicated in Figure 4, the required $\log_{10}$ reduction is decreasing over time for both SARS-CoV-2 and E. coli. However, the decreasing trend for E. coli is much smaller than that of SARS-CoV-2 due to its long persistence period. The results indicate that fabric and leather surfaces contaminated with SARS-CoV-2 need no $\log_{10}$ reduction after 11 hours. A 2 $\log_{10}$ reduction would result in a risk of less than 10^{-6} for leather when the infection risk is relatively high for the first several hours. For fabric, $\log_{10}$ reductions of 1 and 2 are needed to achieve an infection risk of less than 10^{-6} for SARS-CoV-2 and E. coli, respectively. For ceramic, glass, metal, plastic, and polished stone, a 3 $\log_{10}$ reduction is needed to lower the infection risk below 10^{-6} for SARS-CoV-2 and E. coli. Note that paper doesn't require disinfection because of its infection risk of below 10^{-6} for both SARS-CoV-2 and E. coli. To be conservative, a 0.5 $\log_{10}$ reduction of bioburden is used for the paper surfaces.

The disinfection methods used in infrastructure facilities typically consist of UVC light, spray mode, and wipe mode. Each mode has its advantages and disadvantages and selecting the suitable mode depends upon the type and condition of the contaminated surfaces. Table 4 provides the disinfection mode and required $\log_{10}$ reduction to lower the risk below 10^{-6} for different surfaces. Note that $\log_{10}$ reductions for SARS-CoV-2 and E. coli in Table 4 were obtained within the first several hours when the risk was high. The wipe disinfection mode is not suitable for fabric and paper. For paper materials, the spray mode is also not applicable. For leather and hard surfaces, all disinfection methods are considered to be applicable modes. The disinfection level has different implementation methods for each disinfection mode. For UVC light, the disinfection level can be achieved by changing the distance, irradiance level, and exposure time. For wipe mode, the variables include wipe force, contact time, and disinfectant concentration. The spray mode can change the disinfectant concentration and amount to achieve different disinfection levels.

Figure 5 gives an example of the disinfection parameters for plastic and fabric surfaces contaminated by SARS-CoV-2. To reduce the infection risk below 10^{-6}, $\log_{10}$ reductions of 3 and 1 are needed for plastic and fabric, respectively. The UVC light irradiance varies from 100 to 1000 W/m2. The exposure time is set from 1 to 60 seconds. The minimum distance from the UVC light to the contaminated surfaces is 1 cm. The robot can then select suitable disinfection parameters for efficient disinfection.



(a) SARS-CoV-2



(b) E. coli

FIGURE 4. Risk of infection variation over time. (a) SARS-CoV-2; (b) E. coli

TABLE 4. Disinfection mode and level for different materials

Surface type	Material	Disinfection mode			Log_{10} reduction	
		Wipe	Spray	UVC light	SARS-CoV-2	E. coli
Soft	Fabric	-	√	√	1	2
	Leather	√	√	√	2	-
	Paper	-	-	√	0.5	0.5
Hard	Ceramic	√	√	√	3	3
	Glass	√	√	√	3	3
	Metal	√	√	√	3	3
	Plastic	√	√	√	3	3
	Polished stone	√	√	√	3	-
	Wood	√	√	√	3	-

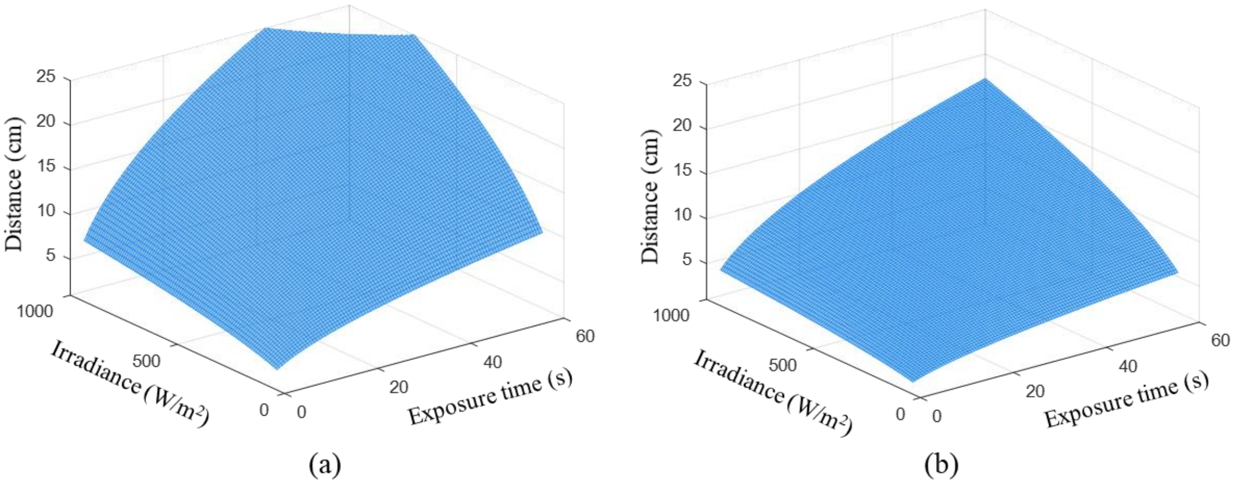


FIGURE 5. UV disinfection parameters for SARS-CoV-2. (a) Fabric; (b) plastic.

4.2 Evaluation on material recognition network

4.2.1 Dataset

The training dataset is prepared using the Materials in Context Database (MINC), which is a large dataset collected from a variety of contexts (Bell, et al., 2015). The MINC contains 2,996,674 single point clicks across 436,749 images, and each click is associated with one of 23 material classes. This study aims to recognize the materials needing disinfection in infrastructure facilities. Some materials like skin and sky in the MINC are not applicable and are discarded for this study. In total, we select 9 types of materials that are commonly seen in infrastructure facilities. These material classes are fabric, leather, paper, ceramic, glass, metal, plastic, polished stone, and wood. To train the CNN model, square image patch data was extracted from the original images. The patch center is defined as the click point, and the size of the patch is 362x362. In many cases, patch areas may go beyond the border of images. Out-of-image pixels were filled with RGB (0,0,0). Note that if the out-of-image pixels number are greater than 262 at any of the two directions, the patch will be removed from the dataset. The patch counts for each material class are shown in Table 5. The training, validation, and test datasets are obtained from the provided train/validation/test splits, which include 1,208,285, 86,228, and 142,857 patches, respectively. Figure 6 presents example patches for each material.

TABLE 5. Sample counts for each material in train, validate and test sets

Material	Train	Validate	test
Fabric	299,929	21,270	36,254
Leather	62,372	4,480	7,313
Paper	17,797	1,242	2,173
Ceramic	21,644	1,571	2,747
Glass	153,492	10,958	17,910
Metal	137,998	9,897	15,850
Plastic	31,282	2,146	3,661
Polished stone	85,196	6,054	9,855
Wood	39,8575	28,610	47,094

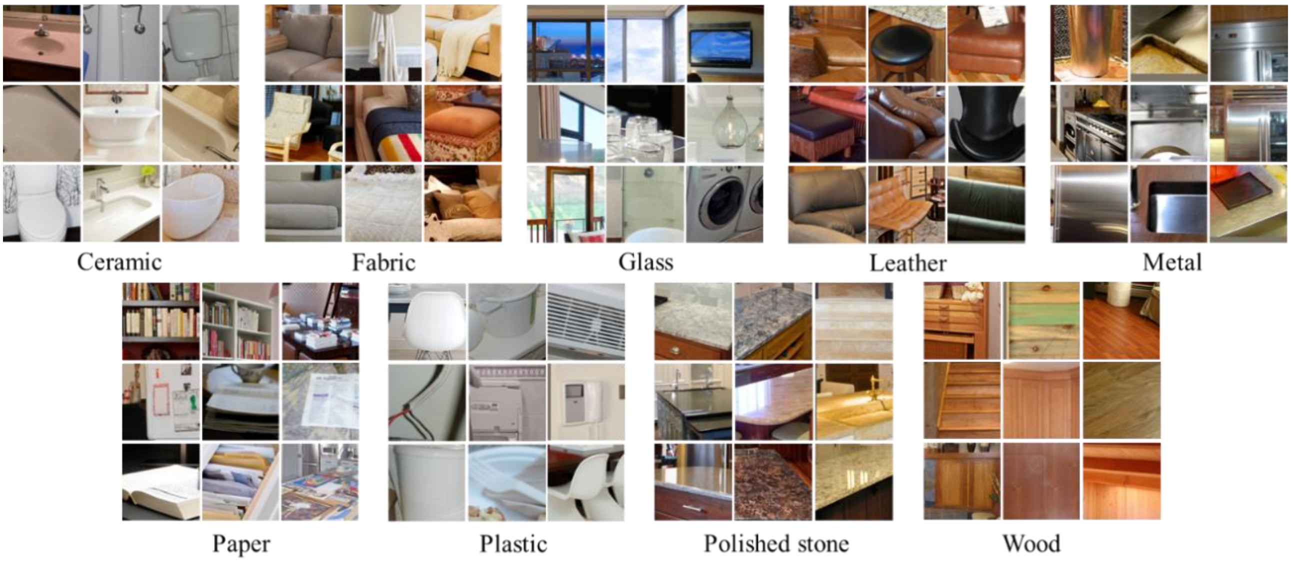


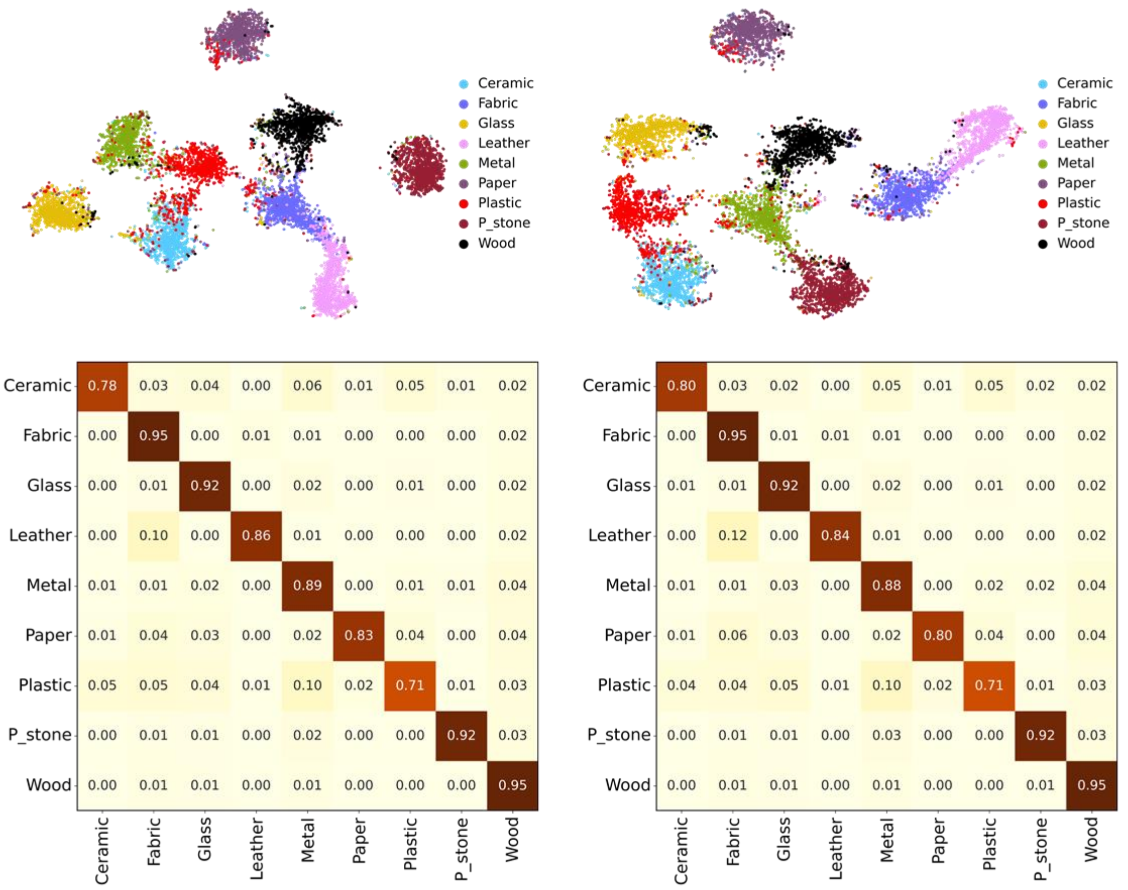
FIGURE 6. Example patches from all 9 material classes with context. Note that the patch center is the associated material (not necessarily the entire patch)

4.2.2 Implementation details

The models were trained using the PyTorch backend (Paszke et al., 2019) with Dual NVIDIA Quadro P5000. The Stochastic Gradient Descent (SGD) optimizer was used with a learning rate of 0.002. The learning rate is divided by 10 for every epoch. The batch size is 128, the weight decay is 0.0001, and the momentum is 0.9. The pretrained weights on ImageNet of the EfficientNet-B4 backbone were used. The model that achieved the best score on the validation dataset is saved and used to further evaluate on the test dataset. Following the precedent set by existing literature (Zhang, et al., 2017; Xue, et al., 2018), the patches are first resized to 256x256. Training samples were augmented by taking random crops measuring 224x224 out of the total 256x256. Horizontal and vertical mirror flips are applied to improve the generalization capability of the network.

4.2.3 Results of material recognition

Figure 7 presents the confusion matrix and Barnes-Hut t-SNE visualization of the material classifications on the validation and test datasets. The trained model achieves high overall accuracy on the validation and test sets, measuring 92.24% and 91.84%, respectively. However, since the validation and test datasets are both imbalanced, recall is a better metric to evaluate the model. Therefore, the confusion matrix is normalized so that diagonal values represent recall for each class. Recall is calculated as the ratio of correctly predicted positives to true positive elements. This is used to measure the model's predictive accuracy for the positive class. The material wood and fabric achieve a high recall score of 95% on both validation and test sets. The glass and polished stone achieve the second-highest score at 92%. A high recall score indicates the predictive power of the trained model on these classes. Plastic has the lowest recall score, 71%, and 10% of plastics are falsely classified as metal. Misclassifying plastic as metal will not cause a change in the disinfection level because both plastic and metal require 3 $\log_{10}$ reductions to reduce infection risk to below 10^{-6}. Furthermore, plastic and metal are both suitable for the wipe, spray, and UVC light disinfection modes. Therefore, plastic classified as metal will not pose a health risk. However, 12% of leathers are misclassified as fabric, which does have an impact on the disinfection level. This is because fabric requires a $\log_{10}$ reduction of 1 to reduce infection risk to target safety level, but a $\log_{10}$ reduction of 2 is required for leather surfaces. As a result, this will potentially cause an incomplete disinfection of the leather surface.



(a) Validation

	Ceramic	Fabric	Glass	Leather	Metal	Paper	Plastic	P_stone	Wood
Ceramic	0.78	0.03	0.04	0.00	0.06	0.01	0.05	0.01	0.02
Fabric	0.00	0.95	0.00	0.01	0.01	0.00	0.00	0.00	0.02
Glass	0.00	0.01	0.92	0.00	0.02	0.00	0.01	0.00	0.02
Leather	0.00	0.10	0.00	0.86	0.01	0.00	0.00	0.00	0.02
Metal	0.01	0.01	0.02	0.00	0.89	0.00	0.01	0.01	0.04
Paper	0.01	0.04	0.03	0.00	0.02	0.83	0.04	0.00	0.04
Plastic	0.05	0.05	0.04	0.01	0.10	0.02	0.71	0.01	0.03
P_stone	0.00	0.01	0.01	0.00	0.02	0.00	0.00	0.92	0.03
Wood	0.00	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.95

(b) Test

	Ceramic	Fabric	Glass	Leather	Metal	Paper	Plastic	P_stone	Wood
Ceramic	0.80	0.03	0.02	0.00	0.05	0.01	0.05	0.02	0.02
Fabric	0.00	0.95	0.01	0.01	0.01	0.00	0.00	0.00	0.02
Glass	0.01	0.01	0.92	0.00	0.02	0.00	0.01	0.00	0.02
Leather	0.00	0.12	0.00	0.84	0.01	0.00	0.00	0.00	0.02
Metal	0.01	0.01	0.03	0.00	0.88	0.00	0.02	0.02	0.04
Paper	0.01	0.06	0.03	0.00	0.02	0.80	0.04	0.00	0.04
Plastic	0.04	0.04	0.05	0.01	0.10	0.02	0.71	0.01	0.03
P_stone	0.00	0.01	0.01	0.00	0.03	0.00	0.00	0.92	0.03
Wood	0.00	0.01	0.01	0.00	0.01	0.00	0.00	0.01	0.95

FIGURE 7. The confusion matrix and Barnes-Hut t-SNE visualization of our method. For confusion matrix, rows are actual classes and columns are predictions; for Barnes-Hut t-SNE, 1110 images for each material class were randomly selected. Each point is associated with an image. The feature map before the classifier was extracted and used for t-SNE visualization

The Barnes-Hut t-SNE (Van der Maaten and Hinton, 2008) algorithm is adopted to visualize CNN feature maps in 2D, which is a tool to visualize high-dimensional features. The t-SNE algorithm assigns a high probability to similar objects and a low probability for dissimilar objects to construct a probability distribution over pairs of high-dimensional data. The T-SNE also constructs a probability distribution over pairs in the low-dimensional map. The location of the points in the map is determined by minimizing the Kullback-Leibler divergence between two distributions. In this study, Principal Component Analysis (PCA) (Wold et al., 1987) is first applied to the feature map to reduce its dimension to 50. The outputs of PCA are then fed into the t-SNE algorithm. The t-SNE algorithm converts the data into a 2D matrix. A relatively large perplexity value of 150 is used in the t-SNE to capture the global structure of the data (Kobak and Berens, 2019).

Figure 7 shows a two-dimensional (2D) representation of the CNN feature map extracted from the layer before the last FC layer. The inputs for the t-SNE visualization are the material class and extracted CNN features. The CNN feature is 1×1 pixel size only, but with 256 channels. These 256 numbers are essentially all the features that the network extracted from the input image. For both validation and test sets, 1110 images are randomly selected from each class for visualization. Each point is associated with an image, and the distance between points approximates the original Euclidean distance in the high-dimensional features. If two image features are similar to each other, they will stay close in the resulting projection in the 2D map. The point color represents its related material classes. The t-SNE plots indicate that points from the same class are organized into clusters, which is an indicator of good differentiation between images of different classes using features before the last fully connected layer. The separated clusters indicate that the proposed network can understand the material data and its classes and is able to differentiate them. The results also highlight the relation between the clusters (i.e., connected clusters indicate there are some semantic relations between materials). For example, the material clusters leather and fabric are connected to each other. These two materials are similar to each other compared to other materials in real life. This is because leather and fabric are both soft materials, and they are popular materials for furniture upholstery. Therefore, it is more challenging for the network to differentiate between these two materials.

To further evaluate the model performance in the context of healthcare facilities, we manually labeled 1,173 patches consisting of 9 material classes. Hospital images were first downloaded from Google Images. These images contained different rooms at hospitals, such as the operating room, the consulting room, the intensive care unit, hallways, day rooms, the ward, and restrooms. We then collected clicks (i.e., single points) in images and assigned material labels to each point. To increase the accuracy of our labels, we only collected clicks with explicit materials. Table 6 presents sample counts for each material. Figure 8 presents example labeled patches for each material in the context of healthcare facilities.

TABLE 6. Sample counts for each material in hospital dataset

Material	Count	Material	Count
Fabric	176	Metal	127
Leather	100	Plastic	100
Paper	100	Polished stone	100
Ceramic	135	Wood	142
Glass	193		

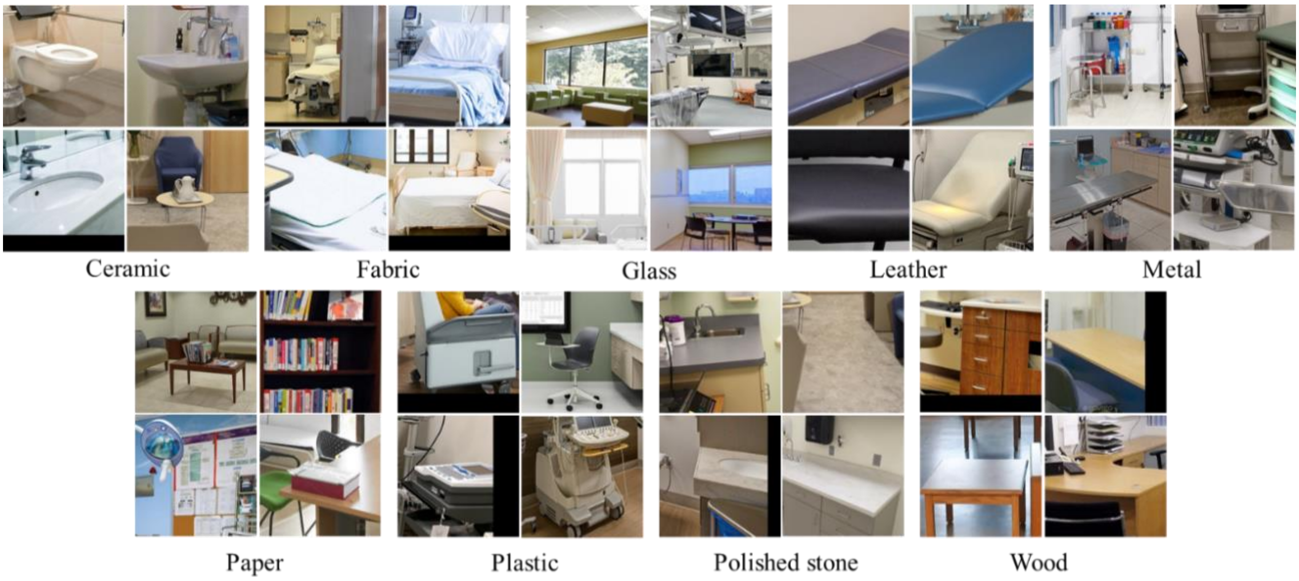
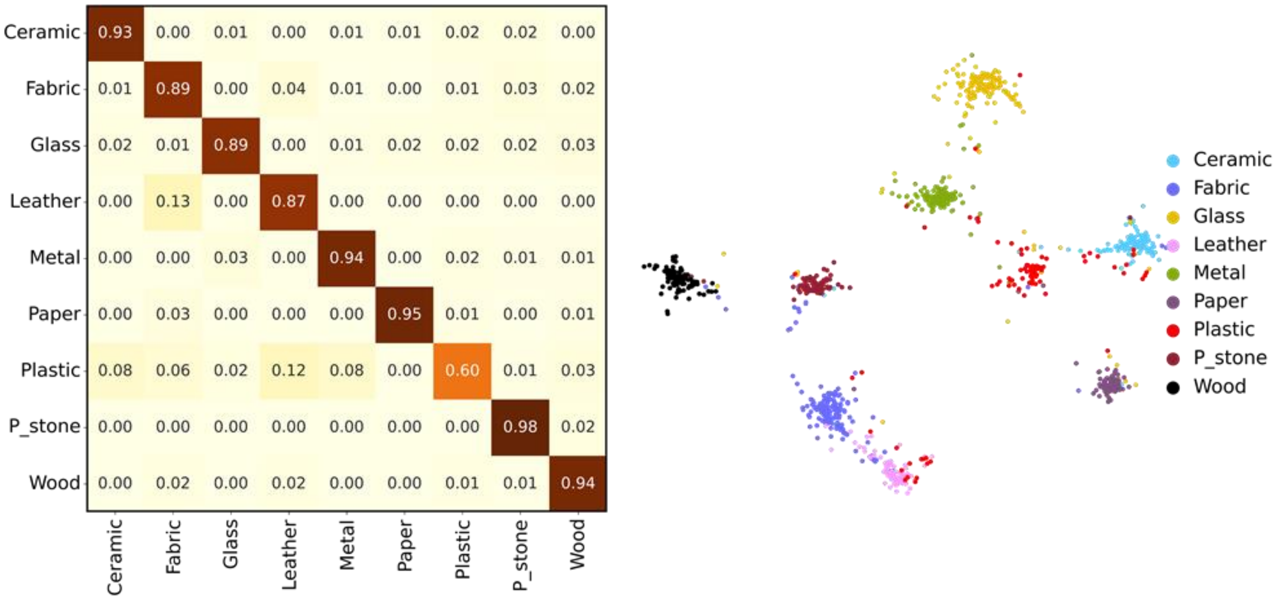


FIGURE 8. Example patches from all 9 material classes at hospitals. Note that the patch center is the associated material

Figure 9 presents the results of the model evaluation on the hospital material dataset. The overall accuracy of the trained model is 89.09%. The confusion matrix indicates that plastic has the smallest recall score at 60%. The recall of plastic is also the smallest on the MINC validation and test sets. This may be attributed to relatively small number of plastic samples in the training set, and the fact that plastic features are similar to other materials like metal and ceramic. The top three false negatives for plastic are leather, ceramic, and metal. Misclassifying plastic as leather will lead to a smaller $\log_{10}$ reduction, which potentially cause inadequate disinfection. The recall score for leather is 87%, which 13% of the leather samples falsely classified as fabric. Other materials achieve high recall scores, demonstrating the efficiency of the trained model. The t-SNE plot shows the same materials are clustered together. The fabric and leather clusters are found to be close to each other. Figure 10 presents correct and incorrect predictions on the hospital material dataset with high confidence.



	Ceramic	Fabric	Glass	Leather	Metal	Paper	Plastic	P_stone	Wood
Ceramic	0.93	0.00	0.01	0.00	0.01	0.01	0.02	0.02	0.00
Fabric	0.01	0.89	0.00	0.04	0.01	0.00	0.01	0.03	0.02
Glass	0.02	0.01	0.89	0.00	0.01	0.02	0.02	0.02	0.03
Leather	0.00	0.13	0.00	0.87	0.00	0.00	0.00	0.00	0.00
Metal	0.00	0.00	0.03	0.00	0.94	0.00	0.02	0.01	0.01
Paper	0.00	0.03	0.00	0.00	0.00	0.95	0.01	0.00	0.01
Plastic	0.08	0.06	0.02	0.12	0.08	0.00	0.60	0.01	0.03
P_stone	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.98	0.02
Wood	0.00	0.02	0.00	0.02	0.00	0.00	0.01	0.01	0.94

(a) Confusion matrix

(b) t-SNE plot

FIGURE 9. The confusion matrix and Barnes-Hut t-SNE visualization of our method on hospital dataset

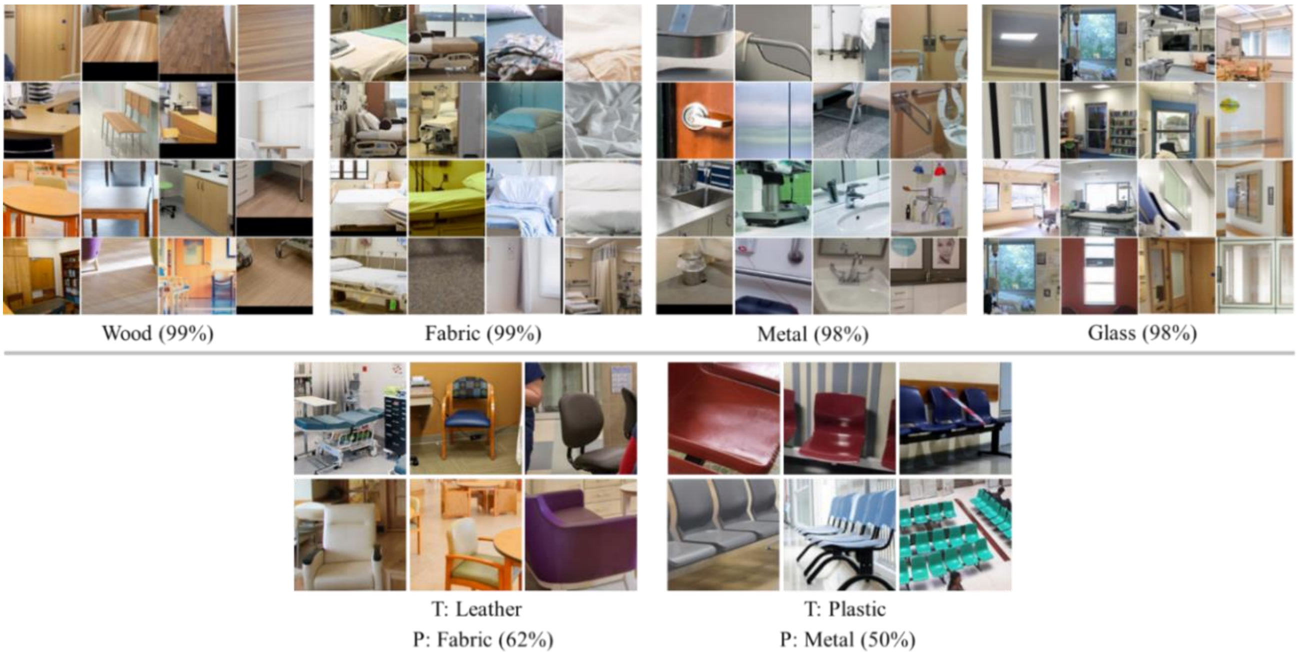


FIGURE 10. Samples with high confidence predictions in hospital material dataset. The first row is correct predictions, and the second row is incorrect predictions (T: actual material, P: predicted). The percentages shown are at least this confident

4.2.4 Comparison to the state-of-the-art methods

We further evaluated the proposed network by comparing it to state-of-the-art methods on six material/texture datasets. These datasets are described as follows. (a) A subset of Material in Context Database (MINC) (Bell, et al., 2015) datasets called MINC-2500 contains 23 material classes and 2,500 images per class. (b) Flicker Material Dataset (FMD) (Sharan et al., 2014) contains a total of 1000 images that are equally distributed across 10 material categories, which has been used as an evaluation benchmark. (c) Ground Terrain in Outdoor Scenes Dataset (GTOS) (Xue, et al., 2018) contains more than 30,000 images with 40 material classes. (d) GTOS-Mobile (Xue, et al., 2018) is a ground terrain dataset captured by mobile phones that consists of 93,945 training images and 6,066 testing images. (e) Describable Textures Database (DTD) (Cimpoi et al., 2014) is a texture database consisting of 5640 images covering 47 classes. Each class consists of 120 images. (f) KTH-TIPS-2b (KTH) (Mallikarjuna et al., 2006) is a material dataset, which is composed of 11 material classes with four samples per class. Each sample contains 108 images.

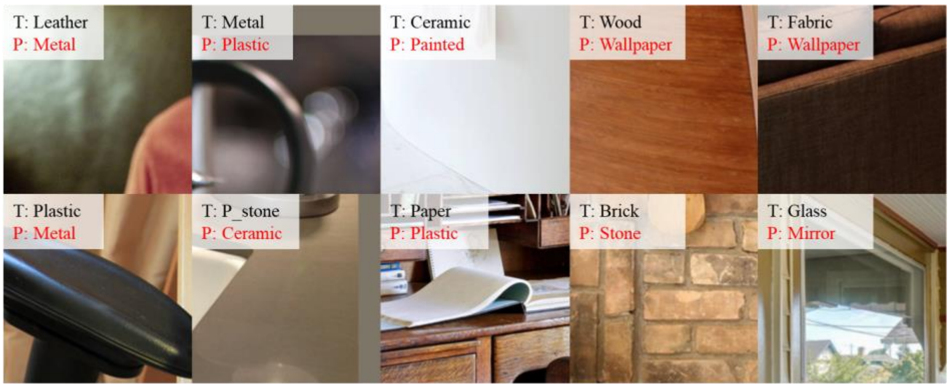
For a fair comparison with other methods (Zhai, et al., 2019; Zhai, et al., 2020; Chen, et al., 2021), the evaluation is based on the provided train-test random splits for MINC-2500, DTD, GTOS-Mobile, and GTOS datasets. As for FMD, the dataset is randomly split into a train-test split in each run with 90 images per class used for training and 10 images used for testing. For KTH, three samples are randomly picked for training and the rest for testing in each run. The results are based on the 5-time run statistics for all the datasets. The learning rate is 0.01 and decays by a factor of 0.1 for every 10 epochs on MINC-2500. For FMD, GTOS, GTOS-Mobile, DTD, and KTH, the learning rate is set to 0.01. The network is trained using momentum of 0.9, weight decay of 0.0001. The training is finished in 30 epochs.

The performance of the proposed method is compared with the Fisher Vector CNN (FV-CN) with a VGG-VD backbone (Cimpoi, et al., 2016), Bilinear-CNN (B-CNN) (Lin and Maji, 2016), Locally-Transferred Fisher Vectors (LFV) (Song, et al., 2017), First and Second-Order information fusion Network (FASON) (Dai et al., 2017), Deep Texture Encoding Network (Deep-TEN) (Zhang, et al., 2017), Deep Encoding Pooling Network (DEP) (Xue, et al., 2018), Deep Multiple-Attribute-Perceived Network (MAPNet) (Zhai, et al., 2019), Multi-level Texture Encoding and Representation Network (MuLTER) (Hu et al., 2019), Deep Structure-Revealed Network (DSRNet) (Zhai, et al., 2020), and Cross-Layer Aggregation of Statistical Self-similarity (CLASSNet) (Chen, et al., 2021). Table 7 presents the comparison results with these state-of-the-art methods. It should be noted that data distribution is different across the six datasets, which leads to varied performance. In addition, the dataset size and the material categories are also different among different datasets, which could also influence the model performance. Our method has the highest accuracy on the six datasets compared to the other texture/material recognition methods. Specifically, our method showed an improvement of 1.9%/2.2%/0.5%/1.4%/0.3%/0.2% in mean accuracy on MINC-2500, FMD, GTOS, GTOS-Mobile, DTD, and KTH compared to state-of-the-art methods, respectively. The comparison with the state-of-the-art methods demonstrates the robustness and accuracy of our methods for the task of material recognition.

The method performance is further compared to the Deep Encoding Pooling Network (DEP), which is the newest model available for our testing. This comparison aims to understand what kind of image features lead to misclassification by DEP, while our method can correctly recognize. The MINC-2500 and DTD are selected for our comparison. Figure 11 shows some example images that are correctly classified by our method and misclassified by the DEP. There are two major challenging in the material/texture classification for these example images. First, some of these images are smooth and featureless. Second, some images show spatially invariant features. Our proposed method can capture both the low-level texture and color information and the high-level semantic information, thus benefitting images with insignificant features. Furthermore, ASPP and the orderless encoder allow the network to learn the spatial repetitive features of material and textures.

TABLE 7. Comparison to state-of-the-art methods on six material/textures dataset

Method	MINC-2500		FMD		GTOS		GTOS-Mobile		DTD		KTH	
	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
FV-CNN (VGG-VD) (2015)	-	-	79.8	1.8	77.1	-	-	-	72.3	1.0	75.4	1.5
B-CNN (2016)	-	-	77.8	1.9	-	-	75.43	-	69.6	0.7	75.1	2.8
LFV (2017)	-	-	82.1	1.9	-	-	-	-	73.8	1.0	82.6	2.6
FASON (2017)	-	-	-	-	-	-	-	-	72.3	0.6	76.5	2.3
Deep-TEN (2017)	81.3	-	80.2	0.9	84.5	2.9	76.12	-	69.6	-	-	-
DEP (2018)	82.0	-	-	-	-	-	-	-	73.2	-	-	-
MAPNet (2019)	-	-	85.2	0.7	84.7	2.2	86.6	1.5	76.1	0.6	84.5	1.3
MuLTER (2019)	82.2	-	-	-	-	-	78.2	-	-	-	-	-
DSRNet (2020)	-	-	86.0	0.8	85.3	2.0	87.0	1.5	77.6	0.6	85.9	1.3
CLASSNet (2021)	84.0	0.6	86.2	0.9	85.6	2.2	85.7	1.4	74.0	0.5	87.7	1.3
Ours	**85.9**	0.4	**88.4**	1.8	**86.1**	0.5	**88.4**	0.6	**77.9**	0.4	**87.9**	2.4



(a) MINC-2500

(b) DTD

FIGURE 11. Example images in MINC-2500 and DTD that are misclassified by DEP while correctly classified by our method. T is the true class; P is the predicted class by DEP

4.2.5 Ablation study

To evaluate the proposed network, we study three components - the multi-level feature integration (ML), ASPP, and encoder (EC) - and summarize their effects on the model performance. The baseline model is generated by removing the integration of the three components, which becomes an EfficientNet-B4 network. The experiment is designed as follows. The effectiveness of each ML, ASPP, and EC component is evaluated by individually integrating them to the baseline model. A combination of either of the two components is also evaluated based on the accuracy metric. The performance is evaluated on the DTD and FMD. The results are listed in Table 8 for comparison. The results reported in the table represent the accuracy in the form of "mean ± s.t.d.%". The experiment results are detailed below.

Multi-level feature integration. In this part, we study the effects of multi-level feature integration, which is proposed to capture the low-level texture and color information and the high-level semantic information. As shown in Table 8, the multi-level feature significantly improves the model performance on DTD and FMD by 3.2% and 6%, respectively (baseline→baseline+ML). We also conduct experiments on a combination of ML with either of the ASPP and EC components. In detail, ML+ASPP and ML+EC combinations are evaluated, and the results indicate an improvement compared with only ML integration on DTD and FMD. The ML has the highest improvement compared to ASPP and EC, which highlights the effectiveness of multi-level features in material representation.

ASPP. The ASPP component is used to capture multi-scale information, which can learn spatial repetitive features in material textures. Compared to the baseline, the integration of ASPP component improves the performance of the network by 1.1% and 3%, respectively (baseline→baseline+ASPP). ASPP+EC is found to be better than baseline+ASPP, which has an improvement of 1.3% and 4% on DTD and FMD, respectively.

Encoder. In this section, we evaluate the effectiveness of the encoder component, which is designed to capture both texture and local spatial information. As shown in table 8, the performance is improved by 1.6% and 5% (baseline→baseline+EC) on DTD and FMD, respectively. This comparison indicates that the EC component can improve model performance. As mentioned above, a combination of EC with ML or ASPP can further improve performance.

TABLE 8. Ablation study on DTD and FMD. 'ML' is multi-level feature integration. 'EC' is the encoder

Model	ML	ASPP	EC	DTD	FMD
Baseline				72.7±0.5	80.3±1.5
	√			75.9±0.3	86.3±1.5
		√		73.8±0.3	83.3±2.1
			√	74.3±0.4	85.3±1.5
	√	√		76.5±0.3	87.0±1.0
	√		√	77.6±0.7	87.6±1.5
		√	√	75.1±0.9	87.3±2.3

Proposed	√	√	√	**77.9**±0.4	**88.4**±1.8

4.3 Implementation

The material-aware disinfection robot is tested in a virtual environment built based on a patient room at a healthcare facility. The room is used to hospitalize patients with COVID-19. The disinfection robot equipped with UV light is used as an illustration. The disinfection robot first moves to potentially contaminated objects requiring disinfection. The images captured by the camera are then be fed into the CNN network to classify the surface materials. The disinfection robot can adapt the disinfection mode and the parameters according to the surface material. Figure 12 shows some example results on the disinfection of the overbed table, a reported for each dataset. door handle, a book, the seat of a chair, a sofa, and a vase in the patient room. The results indicate the proposed material recognition network can recognize the materials of the object surfaces needing disinfection, which can be leveraged to provide the disinfection mode and parameters for the disinfection robot.

Figure 13 presents an implementation of robot disinfection using UVC light with suitable distance, irradiance, and exposure time. Note that the exposure time is only for the contaminated areas approximately under the UVC light reflective shield. For the entire object surfaces, the robotic arm will first identify waypoints to cover contaminated areas, and then plan its trajectory to each waypoint for disinfection.

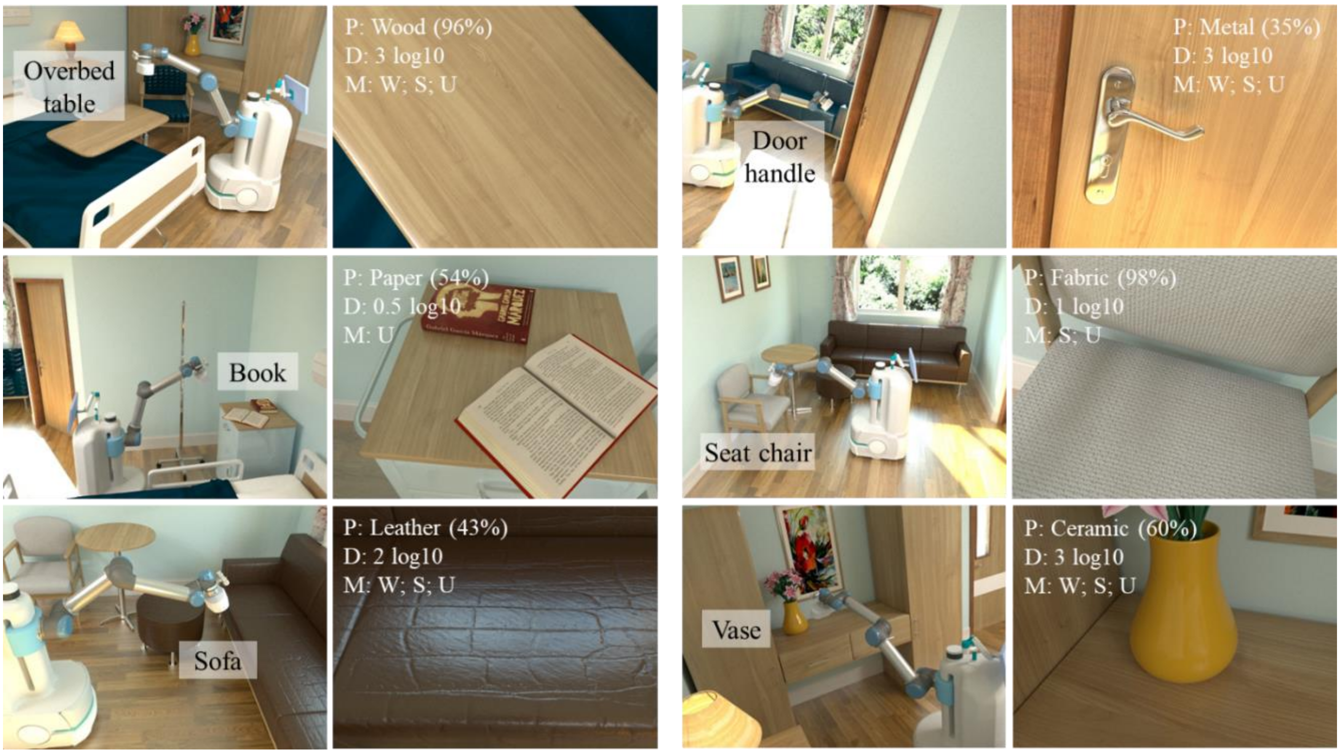


FIGURE 12. Example results of material classification on images captured by the robot. P: predicted material (confidence value in parentheses); D: disinfection dosage; M: disinfection mode; W: wipe; S: spray; U: UVC light

FIGURE 13. Implementation of robot disinfection with suitable parameter

5 DISCUSSIONS

Cleaning and disinfecting object surfaces in infrastructure facilities is critical to mitigating the spread of infectious pathogens and reducing the number of infections. The idea behind this work is to develop methods to recognize surface materials and to computationally link recognized materials with robotic disinfection modes and parameters. To this end, we first developed a new deep learning-based network to recognize the materials of object surfaces needing disinfection. Then, we proposed a computational model to calculate the required $\log_{10}$ reduction for the recognized material and linked it to the applicable robotic disinfection mode and parameters. The developed method has the potential to transform current surface disinfection practices to intelligent robotic disinfection with material-adaptive disinfection parameters. The following sections discuss the robustness of the material recognition method, uncertainty and sensitivity, applicability of robotic disinfection, and limitations and future studies.

5.1 Robustness of material recognition method

This section discusses the potential influence of illumination conditions and the robustness of the proposed approach to illumination. Illumination is related to lighting and weather conditions. Illumination variation is a significant influencing factor for the image classification task. In this experiment, the gamma correction method is adapted to change the illuminance of the image based on the Power-Law Transform function (Poynton, 2012). The image is darker when the gamma values are smaller than 1, and the image is lighter when the gamma values are greater than 1. Gamma values of 0.5, 1.5, 2, 2.5 are investigated in the experiment. Figure 14 presents the material prediction results for wood, fabric, and leather surfaces under varied illumination conditions. The wood surface can be recognized under different illuminations, while the prediction confidence decreases with increasing gamma values. The fabric surface is recognized with high confidence under all investigated illumination conditions. The leather surface is misclassified as fabric when the gamma values are 2 and 2.5. The incorrect predictions stem from the following reasons. First, the prediction confidence is 43% for the original image, which is relatively low compared to wood and fabric. Second, leather surface features are close to fabric surface features, as indicated in Figure 9. The prediction can be improved with more leather material data in various illumination conditions. The illumination conditions have an impact on the performance of our network, particularly for bright images. The network can accurately predict materials in the image under different illumination conditions when the prediction confidence is high for original images. For prediction with a low confidence (e.g., leather), the network also works for slightly brighter and darker conditions. Therefore, our material recognition network can be viewed as robust and reliable regarding illumination variation.

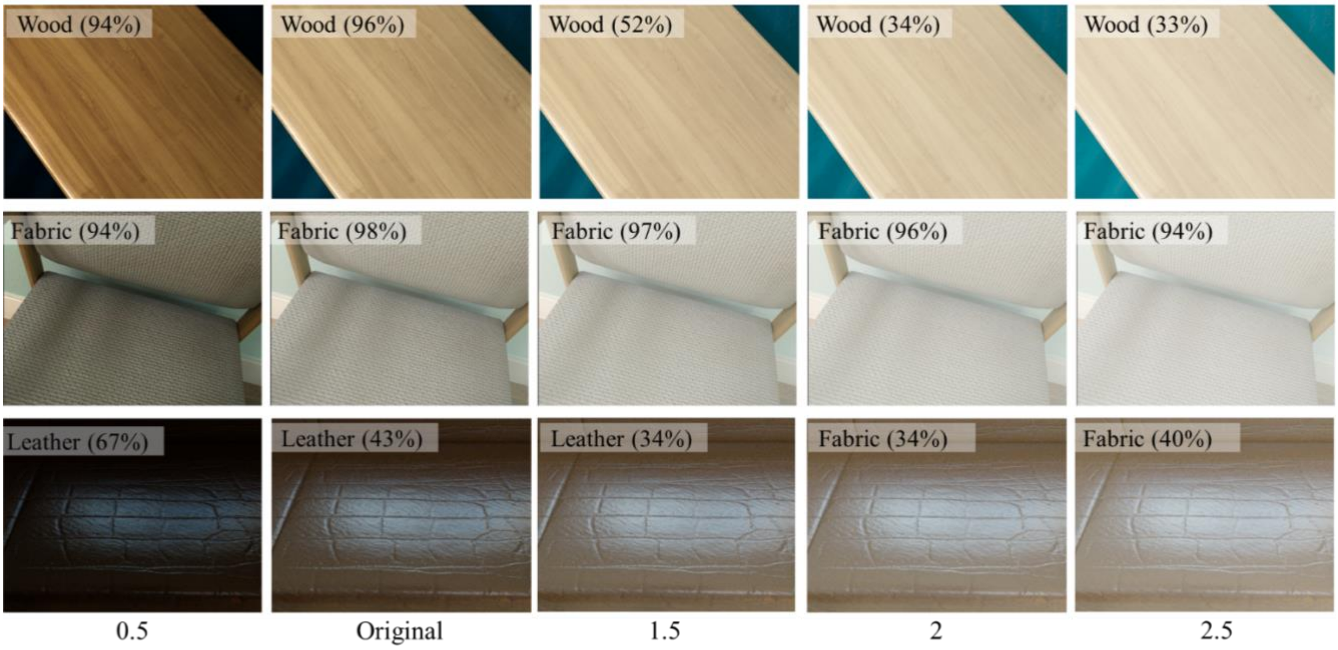


FIGURE 14. Material recognition performance under varied illumination conditions. Predicted material (confidence value in parentheses)

5.2 Uncertainty and sensitivity analysis

In this part, the sensitivity and the uncertainty of the model used to estimate the required $\log_{10}$ reduction are analyzed. Monte Carlo simulations are used to incorporate uncertainty and variability of the input parameters in the risk characterization. A ceramic surface contaminated with SARS-CoV-2 is selected as an illustration. Convergence is tested for the model by running 1000, 5,000, 10,000, 20,000, 50,000, and 100,000 simulations five times. The model estimation becomes stable after 50,000 runs, as indicated in Figure 15. Therefore, our study simulated a total of 50,000 runs for all the models.

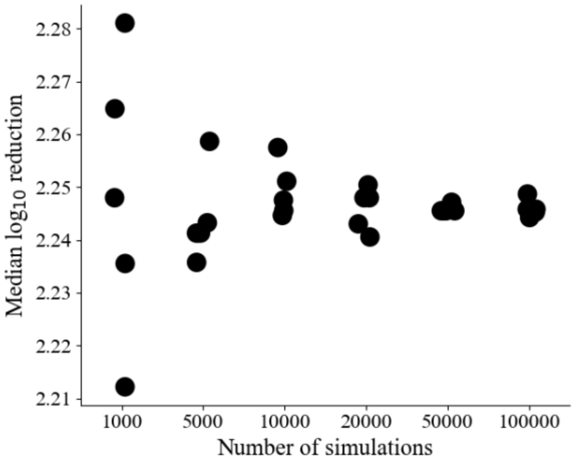


FIGURE 15. Median $\log_{10}$ reduction vs. number of Monte Carlo simulations. The results are based on five runs

The distribution of the disinfection dose for different surfaces is further evaluated with SARS-CoV-2 persisting on the surface after 6 hours. The required $\log_{10}$ reduction distribution is shown in Figure 16. Note that the infection risk for paper is lower than 10^{-6}, which does not require disinfection to control its risk. For fabric and leather, a 2 $\log_{10}$ reduction is found to be sufficient to lower the infection risk. For other materials, the median $\log_{10}$ reduction is between 2 and 3. In some scenarios, required $\log_{10}$ reduction could go beyond 3.

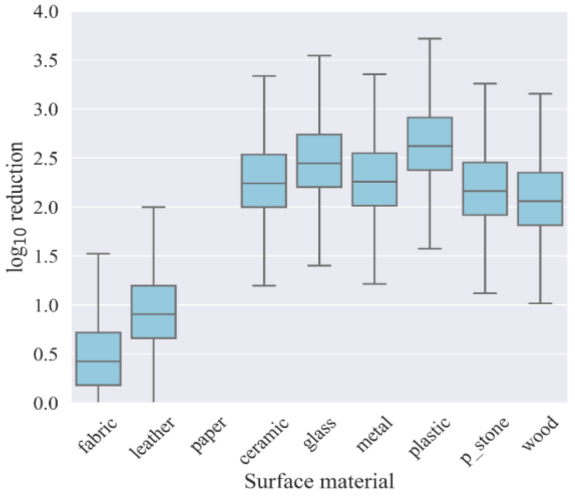


FIGURE 16. Boxplot of $\log_{10}$ reduction for different surface materials contaminated by SARS-CoV-2

Spearman correlation coefficients are used to examine the relationship between the model input parameters and the disinfection dosage. The SARS-CoV-2 is selected for the sensitivity analysis. The transfer efficiency and survival time of SARS-CoV-2 are assumed to be uniformly distributed in the range given in Table 2. In addition, SARS-CoV-2 concentration after 6 hours is used for analysis. A total of 50,000 simulations are conducted. Figure 17 presents the Spearman's correlation coefficients for the input parameters of the model. According to the sensitivity analysis, the model input parameters that mostly influence the required $\log_{10}$ reduction are the transfer efficiency between surface and the hand and the survival time of the pathogen, which are both positively related to the disinfection dosage. These two parameters are material-specific parameters, which further confirms the importance of the material information for disinfection. The pathogen recovery efficiency is negatively correlated with the disinfection dosage. The correlation was positive for all other modeled parameters.

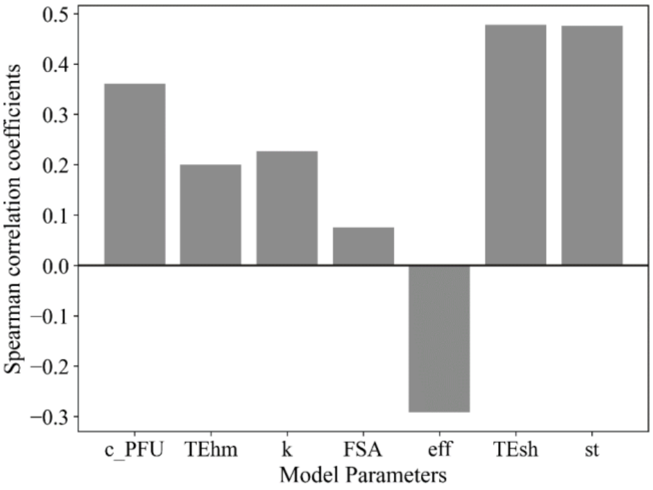


FIGURE 17. Spearman's correlation coefficients for the parameters used in estimating required disinfection dosage. Parameters are abbreviated as follows: c_PFU = conversion factor from Genome copies to the infectious virus in PFU; TEhm = = transfer efficiency of viruses from hand to mucous; k = dose-response parameter; FSA = fractional surface area; eff = pathogen recovery efficiency; TEsh = transfer efficiency of viruses from surface to hand; st = survival time of pathogen

5.3 Applicability of the robotic disinfection

Our proposed robotic disinfection system was successfully implemented in a fully modeled hospital room. The proposed material recognition method achieved an overall accuracy of 89.09% on the dataset collected in the context of healthcare facilities. The processing time for material recognition is around 0.04 seconds for a single image. The processing speed can be enhanced by increasing the batch size in the inference. For example, by setting the inference batch size to 64, a processing time of 0.12 seconds is sufficient to predict all the images. The promising results of our method demonstrated its accuracy and efficiency to provide the material information for the disinfection robot. The recognized materials are computationally linked to robotic disinfection modes and parameters, considering fomite-pathogen-human interactions. Compared to manual disinfection, the adaptive system has the potential to not only ensure complete disinfection of contaminated surfaces, but also to improve the disinfection efficiency by lowering the exposure time and the disinfection dose for low-risk surfaces. The proposed material-adaptive robotic disinfection framework can lead to an intelligent robotic disinfection paradigm that goes well beyond existing systems that are perceived as roaming UV lights for coarse disinfection.

The disinfection robot market was valued at $493 million in 2020 and is expected to reach $3.31 billion by 2026, with a compound annual growth rate (CAGR) of 36.4% (Mordor, 2021). Many disinfection robots have already been deployed in infrastructure facilities like hospitals and schools. However, no disinfection robots to our knowledge possess the material recognition capabilities to adapt disinfection modes and parameters, which largely restricts their disinfection efficiency. Our proposed framework can be integrated with existing disinfection robot platforms to improve their performance. For instance, in Hu et al. (2020a), the UV light wand was used as an end-effector of a robotic arm to disinfect contaminated surfaces in the built environment. With our newly developed framework, UV light parameters can be adapted based on the object surface materials with appropriate exposure time, distance, and irradiance. As such, surfaces can be thoroughly disinfected and free of pathogens in sufficient numbers to prevent disease transmission.

There still exists some obstacles from the concrete operationalization of the proposed robotic disinfection system in the real world. First, controlled experiments need to be conducted to evaluate the effectiveness of the robot by measuring the surface pathogen concentration before and after disinfection. In addition, there still lacks evidence about how much contamination could lead to infection in humans. Second, the disinfection robot needs to be endowed with human activity recognition capability in the future to be able to conduct disinfection tasks in the presence of humans. Finally, there needs to be a validated, reproducible, and documented disinfection protocol for the robot. The development of such a protocol needs to have a close collaboration with the end-users, such as hospitals. As such, the robot design and protocol can be updated based on their feedbacks.

5.4 Limitation and future studies

This study suffers from several limitations that deserve future studies. First, despite the overall high performance of the material recognition network, the performance on plastic surfaces achieved a lower accuracy compared to other material categories. This underperformance was caused by a relatively small number of plastic samples in the training dataset when compared to other materials. In addition, the overall accuracy of the material classification model is smaller for the hospital material dataset than the accuracy for the MINC validation and test sets. The relatively lower accuracy stems from a lack of surface materials collected at hospitals in the training dataset. In the future, more surface material data needs to be collected in the healthcare facilities to fine-tune the network, especially for materials with fewer samples. In addition, other sensory data, such as thermal and time-of-flight depth cameras, could be integrated into the deep learning network to create a more robust model. Second, our work primarily focuses on recognizing materials in context, which does not differentiate the interface between different materials. A semantic segmentation approach is needed to classify materials at a pixel level. However, the segmentation task requires pixel-level annotations, which is expensive and time-consuming. In future research, material segmentation would be an interesting area to explore when more data becomes available.

The third limitation is that this paper investigates a particular pathogen on surfaces to demonstrate the computational feasibility and the complete loop from robotic perception to robotic actions. However, many pathogens can cohabit on the same surfaces in healthcare facilities. In this case, the proposed method needs to be adapted for a multi-pathogen infectious disease system, which requires more research regarding pathogen dependency. Furthermore, for transmission risk, as types and frequencies of human activities and the diversity of environmental surfaces differs between settings, social and environmental contexts are of great importance in assessing the infection risk through fomite transmission. More advanced methods, such as the Environmental Infection Transmission System (EITS) modeling framework proposed in (Li et al., 2009), could be explored to model more complex scenarios. Lastly, the disinfection dose and mode are determined solely based on the surface materials. However, other factors, such as the type of equipment, surface geometry, human touch frequency, and location, are also important considerations for the disinfection practice. Future studies are needed to develop more comprehensive disinfection guidance by integrating these critical factors.

6 CONCLUSIONS

This study proposed a new computational process and deep learning-based material recognition network to classify object surface materials and to adapt disinfection modes and parameters to disinfect surfaces thoroughly and efficiently. The deep learning network integrated multi-level and multi-scale CNN features, as well as a texture encoder to achieve material recognition with high accuracy. The trained network was evaluated on MINC validation and test dataset, and the results achieved an accuracy of 92.24% and 91.84%, respectively. The network achieved an accuracy of 89.09% on a small material dataset containing 1,173 samples collected in the context of healthcare facilities. Furthermore, the proposed material recognition network achieved state-of-the-art results compared to other texture/material recognition methods. The fomite transmission model was adapted to estimate the infection risk for different surfaces and to quantitate the $\log_{10}$ reduction needed to reach the safety target levels. The results indicated that hard surfaces, such as plastic and metal, require a higher disinfection level compared to soft surfaces, such as paper and fabric. The disinfection level was combined with the applicable mode to calculate disinfection parameters for the robot to implement. The adaptive robotic disinfection was successfully implemented in the context of healthcare facilities.

ACKNOWLEDGMENT

This research was funded by the US National Science Foundation (NSF) via Grant numbers: 2026719, 1952140, and 2038967. This research also received support from the Science Alliance at the University of Tennessee Knoxville (UTK) via the Joint Directed Research and Development Program. The authors gratefully acknowledge support from NSF and UTK. Any opinions, findings, recommendations, and conclusions in this paper are those of the authors and do not necessarily reflect the views of NSF and UTK.

REFERENCES

Abrahão, J. S., Sacchetto, L., Rezende, I. M., Rodrigues, R. A. L., Crispim, A. P. C., Moura, C., Mendonça, D. C., Reis, E., Souza, F. & Oliveira, G. F. G. (2021), Detection of Sars-Cov-2 Rna on Public Surfaces in a Densely Populated Urban Area of Brazil: A Potential Tool for Monitoring the Circulation of Infected Patients, *Science of The Total Environment*, 766, 142645.

Abu-Zidan, Y., Nguyen, K., Mendis, P., Setunge, S. & Adeli, H. (2021), Design of a Smart Prefabricated Sanitising Chamber for Covid-19 Using Computational Fluid Dynamics, *Journal of Civil Engineering and Management*, 27(2), 139-148.

Anderson, C. E. & Boehm, A. B. (2021), Transfer Rate of Enveloped and Non-Enveloped Viruses between Fingerpads and Surfaces, *Applied and Environmental Microbiology*.

Arabi, S., Haghighat, A. & Sharma, A. (2020), A Deep‐Learning‐Based Computer Vision Solution for Construction Vehicle Detection, *Computer‐Aided Civil and Infrastructure Engineering*, 35(7), 753-767.

AuYeung, W., Canales, R. A. & Leckie, J. O. (2008), The Fraction of Total Hand Surface Area Involved in Young Children's Outdoor Hand-to-Object Contacts, *Environmental Research*, 108(3), 294-299.

Bell, S., Upchurch, P., Snavely, N. & Bala, K. (2015), Material Recognition in the Wild with the Materials in Context Database, *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE.

Bergeron, C. R., Prussing, C., Boerlin, P., Daignault, D., Dutil, L., Reid-Smith, R. J., Zhanel, G. G. & Manges, A. R. (2012), Chicken as Reservoir for Extraintestinal Pathogenic Escherichia Coli in Humans, Canada, *Emerging infectious diseases*, 18(3), 415.

Boehm, A. B. (2019), Risk-Based Water Quality Thresholds for Coliphages in Surface Waters: Effect of Temperature and Contamination Aging, *Environmental Science: Processes & Impacts*, 21(12), 2031-2041.

Bouri, N. & Shatalov, V. (2020), Uvgi Scientific Calculator, *arXiv*.

Caputo, B., Hayman, E. & Mallikarjuna, P. (2005), Class-Specific Material Categorisation, *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, IEEE.

Castaño, N., Cordts, S. C., Kurosu Jalil, M., Zhang, K. S., Koppaka, S., Bick, A. D., Paul, R. & Tang, S. K. Y. (2021), Fomite Transmission, Physicochemical Origin of Virus–Surface Interactions, and Disinfection Strategies for Enveloped Viruses with Applications to Sars-Cov-2, *ACS Omega*, 6(10), 6509-6527.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F. & Adam, H. (2018), Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation, *Computer Vision – ECCV 2018*, Springer International Publishing, pp. 833-851.

Chen, Z., Li, F., Quan, Y., Xu, Y. & Ji, H. (2021), Deep Texture Recognition Via Exploiting Cross-Layer Statistical Self-Similarity, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5231-5240.

Chin, A. W. H., Chu, J. T. S., Perera, M. R. A., Hui, K. P. Y., Yen, H.-L., Chan, M. C. W., Peiris, M. & Poon, L. L. M. (2020), Stability of Sars-Cov-2 in Different Environmental Conditions, *The Lancet Microbe*, 1(1), e10.

Choi, H., Chatterjee, P., Lichtfouse, E., Martel, J. A., Hwang, M., Jinadatha, C. & Sharma, V. K. (2021), Classical and Alternative Disinfection Strategies to Control the Covid-19 Virus in Healthcare Facilities: A Review, *Environmental Chemistry Letters*, 1-7.

Cimpoi, M., Maji, S., Kokkinos, I., Mohamed, S. & Vedaldi, A. (2014), Describing Textures in the Wild, *2014 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE.

Cimpoi, M., Maji, S., Kokkinos, I. & Vedaldi, A. (2016), Deep Filter Banks for Texture Recognition, Description, and Segmentation, *International Journal of Computer Vision*, 118(1), 65-94.

Cinar, A. & OnbaŞI, E. (2021), Monitoring Environmental Microbiological Safety in a Frozen Fruit and Vegetable Plant, *Food Science and Technology*, 41(1), 232-237.

Collivignarelli, M. C., Abbà, A., Benigna, I., Sorlini, S. & Torretta, V. (2018), Overview of the Main Disinfection Processes for Wastewater and Drinking Water Treatment Plants, *Sustainability*, 10(1), 86.

Dai, X., Ng, J. Y.-H. & Davis, L. S. (2017), Fason: First and Second Order Information Fusion Network for Texture Recognition, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE.

Diab-El Schahawi, M., Zingg, W., Vos, M., Humphreys, H., Lopez-Cerero, L., Fueszl, A., Zahar, J. R. & Presterl, E. (2021), Ultraviolet Disinfection Robots to Improve Hospital Cleaning: Real Promise or Just a Gimmick?, *Antimicrobial Resistance & Infection Control*, 10(1), 1-3.

Doll, M., Stevens, M. & Bearman, G. (2018), Environmental Cleaning and Disinfection of Patient Areas, *International Journal of Infectious Diseases*, 67, 52-57.

Dong, E., Du, H. & Gardner, L. (2020), An Interactive Web-Based Dashboard to Track Covid-19 in Real Time, *The Lancet Infectious Diseases*, 20(5), 533-534.

Donskey, C. J. (2013), Does Improving Surface Cleaning and Disinfection Reduce Health Care-Associated Infections?, *American Journal of Infection Control*, 41(5), S12-S19.

Esteves, D. C., Pereira, V. C., Souza, J. M., Keller, R., Simões, R. D., Winkelstroter Eller, L. K. & Rodrigues, M. V. P. (2016), Influence of Biological Fluids in Bacterial Viability on Different Hospital Surfaces and Fomites, *American Journal of Infection Control*, 44(3), 311-314.

Fleming, M., Patrick, A., Gryskevicz, M., Masroor, N., Hassmer, L., Shimp, K., Cooper, K., Doll, M., Stevens, M. & Bearman, G. (2018), Deployment of a Touchless Ultraviolet Light Robot for Terminal Room Disinfection: The Importance of Audit and Feedback, *American Journal of Infection Control*, 46(2), 241-243.

Gibson, D., Kendrick, S., Simpson, E., Costello, D., Davis, R., Szetela, A., McCreary, M. & Schriber, J. (2017), Implementation of Xenon Ultraviolet-C Disinfection Robot to Reduce Hospital Acquired Infections in Hematopoietic Stem Cell Transplant Population, *Biology of Blood and Marrow Transplantation*, 23(3), S472.

Gordon, R. (2020). "Csail Robot Disinfects Greater Boston Food Bank." Retrieved November 8, 2021, from https://news.mit.edu/2020/csail-robot-disinfects-greater-boston-food-bank-covid-19-0629.

Guo, L., Yang, Z., Guo, L., Chen, L., Cheng, Z., Zhang, L. & Long, E. (2021), Study on the Decay Characteristics and Transmission Risk of Respiratory Viruses on the Surface of Objects, *Environmental Research*, 194, 110716.

Haque, M., Sartelli, M., McKimm, J. & Bakar, M. A. (2018), Health Care-Associated Infections–an Overview, *Infection and Drug Resistance*, 11, 2321.

Harvey, A. P., Fuhrmeister, E. R., Cantrell, M. E., Pitol, A. K., Swarthout, J. M., Powers, J. E., Nadimpalli, M. L., Julian, T. R., Pickering, A. J. J. E. S. & Letters, T. (2020), Longitudinal Monitoring of Sars-Cov-2 Rna on High-Touch Surfaces in a Community Setting, *Environmental Science & Technology Letters*, 8(2), 168-175.

Hsieh, Y. A., Yang, Z. & James Tsai, Y. C. (2021), Convolutional Neural Network for Automated Classification of Jointed Plain Concrete Pavement Conditions, *Computer ‐ Aided Civil and Infrastructure Engineering*, 36(11), 1382-1397.

Hu, D., Zhong, H., Li, S., Tan, J. & He, Q. (2020a), Segmenting Areas of Potential Contamination for Adaptive Robotic Disinfection in Built Environments, *Building and Environment*, 184, 107226.

Hu, J., Shen, L., Albanie, S., Sun, G. & Wu, E. (2020b), Squeeze-and-Excitation Networks, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(8), 2011-2023.

Hu, Y., Long, Z. & AlRegib, G. (2019), Multi-Level Texture Encoding and Representation (Multer) Based on Deep Neural Networks, *2019 IEEE International Conference on Image Processing (ICIP)*, IEEE.

Kang, J., Park, Y.-J., Lee, J., Wang, S.-H. & Eom, D.-S. (2018), Novel Leakage Detection by Ensemble Cnn-Svm and Graph-Based Localization in Water Distribution Systems, *IEEE Transactions on Industrial Electronics*, 65(5), 4279-4289.

Kobak, D. & Berens, P. (2019), The Art of Using T-Sne for Single-Cell Transcriptomics, *Nature Communications*, 10(1).

Kowalski, W. (2009), Mathematical Modeling of Uv Disinfection, *Ultraviolet Germicidal Irradiation Handbook*, Springer, pp. 51-72.

Kraay, A. N. M., Hayashi, M. A. L., Hernandez-Ceron, N., Spicknall, I. H., Eisenberg, M. C., Meza, R. & Eisenberg, J. N. S. (2018), Fomite-Mediated Transmission as a Sufficient Pathway: A Comparative Analysis across Three Viral Pathogens, *BMC Infectious Diseases*, 18(1).

Leung, N. H. (2021), Transmissibility and Transmission of Respiratory Viruses, *Nature Reviews Microbiology*, 1-18.

Lewis, D. (2021), Why Indoor Spaces Are Still Prime Covid Hotspots, *Nature*, 592(7852), 22-25.

Li, S., Eisenberg, J. N., Spicknall, I. H. & Koopman, J. S. (2009), Dynamics and Control of Infections Transmitted from Person to Person through the Environment, *American journal of epidemiology*, 170(2), 257-265.

Lin, T.-Y. & Maji, S. (2016), Visualizing and Understanding Deep Texture Representations, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE.

Lopez, G. U., Gerba, C. P., Tamimi, A. H., Kitajima, M., Maxwell, S. L. & Rose, J. B. (2013), Transfer Efficiency of Bacteria and Viruses from Porous and Nonporous Fomites to Fingers under Different Relative Humidity Conditions, *Applied and Environmental Microbiology*, 79(18), 5728-5734.

Luo, X., Li, H., Yu, Y., Zhou, C. & Cao, D. (2020), Combining Deep Features and Activity Context to Improve Recognition of Activities of Workers in Groups, *Computer ‐ Aided Civil and Infrastructure Engineering*, 35(9), 965-978.

Mallikarjuna, P., Targhi, A. T., Fritz, M., Hayman, E., Caputo, B. & Eklundh, J.-O. (2006), The Kth-Tips2 Database, Computational Vision Active Perception Laboratory, Stockholm, Sweden, pp. 1-10.

McGinn, C., Scott, R., Donnelly, N., Roberts, K. L., Bogue, M., Kiernan, C. & Beckett, M. (2021), Exploring the Applicability of Robot-Assisted Uv Disinfection in Radiology, *Frontiers in Robotics and AI*, 7.

Mordor, I. (2021). "Disinfectant Robot Market - Growth, Trends, Covid-19 Impact, and Forecasts (2021 - 2026)." Retrieved November 8, 2021, from https://www.mordorintelligence.com/industry-reports/disinfectant-robot-market.

Mudgal, P., Breidt, F., Lubkin, S. R. & Sandeep, K. P. (2006), Quantifying the Significance of Phage Attack on Starter Cultures: A Mechanistic Model for Population Dynamics of Phage and Their Hosts Isolated from Fermenting Sauerkraut, *Applied and Environmental Microbiology*, 72(6), 3908-3915.

NOVUS (2021). "Intelligent Robots for Targeted Combating of Viruses and Bacteria." Retrieved October 4, 2021, from https://www.novuslight.com/intelligent-robots-for-targeted-combating-of-viruses-and-bacteria_N11323.html.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N. & Antiga, L. (2019), Pytorch: An Imperative Style, High-Performance Deep Learning Library, *Advances in neural information processing systems*, 32, 8026-8037.

Pitol, A. K., Bischel, H. N., Kohn, T. & Julian, T. R. (2017), Virus Transfer at the Skin–Liquid Interface, *Environmental Science & Technology*, 51(24), 14417-14425.

Pitol, A. K. & Julian, T. R. (2021), Community Transmission of Sars-Cov-2 by Surfaces: Risks and Risk Reduction Strategies, *Environmental Science & Technology Letters*, 8(3), 263-269.

Poynton, C. (2012), *Digital Video and Hd: Algorithms and Interfaces*, Elsevier.

Rafiei, M. H. & Adeli, H. (2017), Neews: A Novel Earthquake Early Warning Model Using Neural Dynamic Classification and Neural Dynamic Optimization, *Soil Dynamics and Earthquake Engineering*, 100, 417-427.

Ramalingam, B., Yin, J., Rajesh Elara, M., Tamilselvam, Y. K., Mohan Rayguru, M., Muthugala, M. & Félix Gómez, B. (2020), A Human Support Robot for the Cleaning and Maintenance of Door Handles Using a Deep-Learning Framework, *Sensors*, 20(12), 3543.

Richter, W. R., Sunderman, M. M., Wendling, M. Q. S., Serre, S., Mickelsen, L., Rupert, R., Wood, J., Choi, Y., Willenberg, Z. & Calfee, M. W. (2019), Evaluation of Altered Environmental Conditions as a Decontamination Approach for Nonspore ‐ Forming Biological Agents, *Journal of Applied Microbiology*, 128(4), 1050-1059.

Rutala, W. A. & Weber, D. J. (2016), Monitoring and Improving the Effectiveness of Surface Cleaning and Disinfection, *American Journal of Infection Control*, 44(5), e69-e76.

Sharan, L., Rosenholtz, R. & Adelson, E. H. (2014), Accuracy and Speed of Material Categorization in Real-World Images, *Journal of Vision*, 14(9), 12-12.

Song, Y., Zhang, F., Li, Q., Huang, H., O'Donnell, L. J. & Cai, W. (2017), Locally-Transferred Fisher Vectors for Texture Classification, *2017 IEEE International Conference on Computer Vision (ICCV)*, IEEE.

Tan, M. & Le, Q. (2019), Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks, *International Conference on Machine Learning*, PMLR, pp. 6105-6114.

Thakar, S., Malhan, R. K., Bhatt, P. M. & Gupta, S. K. (2021), Area-Coverage Planning for Spray-Based Surface Disinfection with a Mobile Manipulator, *Robotics and Autonomous Systems*, 103920.

Trindade, S. N. C., Pinheiro, J. S., de Almeida, H. G., Pereira, K. C. & Sobrinho, P. d. S. C. (2014), Bacteriological Quality and Food Safety in a Brazilian School Food Program, *Nutricion hospitalaria*, 29(1), 80-87.

Van der Maaten, L. & Hinton, G. (2008), Visualizing Data Using T-Sne, *Journal of machine learning research*, 9(11).

Vincent, J. (2020). "Toyota's Robot Butler Prototype Hangs from the Ceiling Like a Bat." Retrieved October 4, 2021, from https://www.theverge.com/2020/10/1/21496692/toyota-robots-tri-research-institute-home-helping-gantry-ceiling-machine.

Virtanen, J., Aaltonen, K., Kivistö, I. & Sironen, T. (2021), Survival of Sars-Cov-2 on Clothing Materials, *Advances in Virology*, 2021, 1-5.

Warnes, S. L., Little, Z. R. & Keevil, C. W. (2015), Human Coronavirus 229e Remains Infectious on Common Touch Surface Materials, *mBio*, 6(6).

Wißmann, J. E., Kirchhoff, L., Brüggemann, Y., Todt, D., Steinmann, J. & Steinmann, E. (2021), Persistence of Pathogens on Inanimate Surfaces: A Narrative Review, *Microorganisms*, 9(2), 343.

Wold, S., Esbensen, K. & Geladi, P. (1987), Principal Component Analysis, *Chemometrics and Intelligent Laboratory Systems*, 2(1-3), 37-52.

Xue, J., Zhang, H. & Dana, K. (2018), Deep Texture Manifold for Ground Terrain Recognition, *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE.

Xue, J., Zhang, H., Dana, K. & Nishino, K. (2017), Differential Angular Imaging for Material Recognition, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE.

Zemmar, A., Lozano, A. M. & Nelson, B. J. (2020), The Rise of Robots in Surgical Environments During Covid-19, *Nature Machine Intelligence*, 2(10), 566-572.

Zhai, W., Cao, Y., Zha, Z.-J., Xie, H. & Wu, F. (2020), Deep Structure-Revealed Network for Texture Recognition, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11010-11019.

Zhai, W., Cao, Y., Zhang, J. & Zha, Z.-J. (2019), Deep Multiple-Attribute-Perceived Network for Real-World Texture Recognition, *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE.

Zhang, H., Xue, J. & Dana, K. (2017), Deep Ten: Texture Encoding Network, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE.

Zhao, Y.-L., Huang, H.-P., Chen, T.-L., Chiang, P.-C., Chen, Y.-H., Yeh, J.-H., Huang, C.-H., Lin, J.-F. & Weng, W.-T. (2021), A Smart Sterilization Robot System with Chlorine Dioxide for Spray Disinfection, *IEEE Sensors Journal*.