ABSTRACT
Recognizing materials using image analysis is a classic problem. However, little research has been done with the images which have visual impediments such as noise, obstacle, or painting. This paper introduces the problem of recognizing covered materials which are distorted visually (e.g., materials covered by graffiti). We propose a set of approaches to solve this problem using a class of deep learning and transfer learning models, and evaluate our approaches empirically using a large-scale real world dataset that displays street scenes containing various materials which are covered with graffiti. Our experiments show that recognizing covered materials using the state-of-the-art approach for material recognition produced an mAP of 19%, while our proposed approach achieved an mAP of 60%. This evidently demonstrates that an approach for plain material recognition is not suitable for recognizing covered materials; hence this problem should be treated differently as in our proposed approaches.

Index Terms— Surface Material, Covered Material Recognition, Material Classification, Graffiti

1. INTRODUCTION
Recently, cities are getting more equipped with technologies to improve their operational efficiency and quality of life of their inhabitants, transforming themselves into smart cities. The ubiquity of camera sensors (e.g., surveillance camera, smartphones, and UAVs) enables the availability of rich visual data about city streets; hence developing various image-based smart city applications [1, 2, 3, 4, 5, 6, 7, 8] (e.g., street cleanliness classification [1], road damage identification [2, 3], traffic flow analysis [4], and situation awareness of disasters [5]). One application is to understand the materials of objects in a scene. Material recognition has been widely investigated [9, 10, 11, 12, 13, 14]. However, sometimes, materials are covered by incidental “covers” (referred to as covered materials) such as graffiti drawn by people or property damages caused by natural disasters. These covers distort the visual characteristics of the underlying materials; hence make the recognition of covered materials challenging.

Nonetheless, covered material recognition is essential for developing efficient management solutions in smart cities (e.g., recognizing materials of damaged properties for post-disaster management, recognizing covered materials by graffiti for removal planning, and accurate estimation of building energy consumption).

Technically, covered material recognition comprises two sub-tasks: detecting a cover as well as recognizing its underlying material. Utilizing a special device, such as hyperspectral imaging cameras (based on measuring reflectance), for material recognition [15] can solve the task of covered material recognition accurately; however, it is expensive and impractical in large scale monitoring. Alternatively, advances in computer vision enable using user-generated images from regular cameras. However, the state-of-the-art image-based approaches [13, 14] for plain material recognition may be inefficient for the covered material recognition problem. Therefore, we propose a class of learning approaches. The first approach generates a learned model assuming a cover and its underlying material as a unified object. The second approach cascades two models (cover detection and material recognition) to tackle the two sub-tasks of the problem separately. These two approaches are enhanced utilizing some heuristics. To evaluate our approaches, this paper focuses on recognizing materials covered by graffiti as a case study using a big real image dataset from the Los Angeles Sanitation Department (LASAN). The source code of our approaches is available at (https://github.com/dweeptrivedi/covered-material-recognition).

2. RELATED WORK
Material Dataset & Recognition. Various material datasets such as Flickr Material Dataset (FMD) [19] and Material in Context Database (MINC) [13] have been available. To address the material recognition problem, researchers have used image-based classification techniques using various classical image feature vectors (e.g., the approach proposed by Liu et al. [9] use bag-of-words consisting of color, SIFT, and

1Such images can be collected by spatial crowdsourcing mechanisms [16]. User-generated images are usually tagged with their locations (if not, they can be localized [17]). Hence, to measure if the available visual data is sufficient for analyzing a geographical region, spatial coverage models [18] can be used.
reflectance-based edge features, Hu et al. [10] uses variances of oriented features, and Qi et al. [11] uses pairwise local binary pattern features). Thereafter, due to the advances of convolutional neural networks, Cimpoi et al. [12] utilized CNN to develop an improved Fisher vector. Bell et al. [13] used the transfer learning mechanism by fine-tuning a pre-trained model (GoogLeNet [20]) on image segments from the MINC dataset while the GoogLeNet model was originally trained for the object recognition task.

**Graffiti Detection & Retrieval.** Many image-based research works have been devoted to graffiti detection and retrieval. Some of them focused on developing algorithms to automatically detect the act of graffiti drawing on a surface using surveillance cameras based on motion changes [21], visual changes due to both brightness and depth [22], or geometry changes [23]. Furthermore, various image-based research systems have been developed for detecting graffiti such as GARI [24], Graffiti Tracker [25], and GRIP [26]. Another body of research work focused on the retrieval of graffiti images from an image database following various paradigms: a) similarity-based retrieval using the SIFT features [27], b) semantic-based retrieval using optical character recognition (OCR) techniques [28], and c) author-based retrieval by analyzing the graffiti style (e.g., shape context and word matching) [29]. However, to the best of our knowledge, there is no existing work focusing on both graffiti detection and underlying material recognition.

### 3. PROBLEM DEFINITION AND DATASET

**Definition 1 (Covered Material).** A covered material is a material \( m \) that is covered partially with a cover \( c \) that distorts the visual appearance of \( m \).

In nature, materials can be categorized into various types. For example, FMD [19] uses a material classification of 10 types (i.e., fabric, foliage, glass, leather, metal, paper, plastic, stone, water, and wood) while the MINC classification [13] consists of 23 types. In our work, we consider only 5 types of materials which can be easily found on urban streets: stone, wood, metal, glass, and fabric. One important observation in recognizing materials in real-world scenes is that they may be covered, visually by various covers such as graffiti and damages. Such visual covers distort the characteristics of their underlying materials, and hence make their recognition challenging.

**Definition 2 (Covered Material Recognition).** Given an image \( I \), a certain type of cover \( c_k \), and \( t \) types of materials (i.e., \( M = \{m_1, m_2, \ldots, m_t\} \)), the covered material recognition is to detect the region in \( I \) which contains a cover of type \( c_k \) and to classify the type of its underlying material \( m_j \in M \).

The covered material recognition can be used for many applications (e.g., energy consumption calculation and inspection of buildings). Our case study is on graffiti removal application to prepare cleanup equipment for the removal process which requires detailed information about the covered material and object before physically arriving on the site. For example, the graffiti removal on a tree is quite different from that on a metal door. Therefore, our approach considers some common types of street objects\(^2\). In particular, given \( t \) types of materials (i.e., \( M = \{m_1, m_2, \ldots, m_t\} \)) and \( n \) types of objects (i.e., \( O = \{o_1, o_2, \ldots, o_n\} \)), the extended material classification includes at maximum \( t \times n \) classes.

In this case study, we consider 9 common object types on streets (i.e., box, tree, wall, door, pole, fence, window, stairs, and fire hydrant) along with the previously mentioned 5 material types. In our case study, we focus on 18 classes of the extended material classification since many of the \( 5 \times 9 \) classes may be discarded because of being unreasonable (e.g., metal tree) or due to the lack of example images. Some image examples are shown in Fig. 1.

Designing an image-based algorithm for the covered material recognition requires an image dataset as described in Definition 3. The annotations of \( \mathcal{D} \) follow the extended materials classification.

**Definition 3 (Covered Material Image Dataset).** The dataset \( \mathcal{D} \) is composed of \( k \) images (\( \mathcal{D} = \{I_1, I_2, \ldots, I_k\} \)) and each image is annotated with at least one region which displays a covered material of type \( m_j \in M \).

### 4. PROPOSED APPROACHES

Recognizing covered materials is different from the plain material recognition problem. In particular, the covered material recognition is associated with two challenges. First, the visual characteristics of the underlying surface material are partially hidden or distorted due to the existence of a cover. Second, it implies learning the visual characteristics of two distinctive units: cover and underlying material. To address these challenges, we propose the following approaches.

#### 4.1. Straightforward Learning Approach

One straightforward approach is to consider a cover with its corresponding material together as one unified object (referred to as one-phase learning approach (OLA)). Then, each type of unified objects is treated as a unique one. Subsequently, one of the state-of-the-art object detection algorithms (e.g., YOLO [31]) can be trained to learn the combined visual features of both the cover and its underlying material for various image regions displaying covered materials (see Fig. 2).

\(^2\)It was shown by Hu et al. [10] and Zheng et al. [30] that material recognition can be improved by incorporating both material and object recognition.
is distinct from the others. Therefore, the learning approach is expanded to the segment which contains the region of a covered material. The selected segment potentially conveys similar visual characteristics of the covered material; hence enhancing the learning approach. Both expansion methods can be employed with OLA (referred to as OLA with Proportional Expansion (OLA − PE) and OLA with Semantic Expansion (OLA − SE). Similarly, these methods are used with TLA too (i.e., TLA − PE and TLA − SE).

5. EXPERIMENTS

5.1. Dataset, Settings, and Baseline Approach

Dataset. Our dataset consists of 19,000 graffiti images (□) collected by LASAN. Every image was tagged with one or more ground-truth boxes where each box is the minimum rectangle surrounding a graffiti and annotated with one of the 18 types of covered materials. These images were correctly labeled by LASAN experts based on the classification of covered materials to automate the graffiti removal process. The distribution of images among the classes of covered materials is shown in Table 1. Note that the image distribution is imbalanced and hence it may result in a biased trained model. Thus, with sampling mechanisms, additional synthesized images have been generated and added to overcome the imbalance problem. In particular, we used the Python Augmentor library [33] which uses some image processing techniques (e.g., sharpen, affine transformation, rotation, and contrast normalization) for the categories with low frequency.

Implementation Settings. To implement our approaches, we created a detection model (for OLA and the first model of TLA) by fine-tuning the darknet53 model using the YOLOv3 framework (specifically, using stochastic gradient descent (SGD) with a batch size of 128 and learning rate of 0.001). For implementing the classification model in the second phase of TLA, we used the Caffe architecture [34] to fine-tune the GoogLeNet model [20] using our dataset (using SGD with a batch size of 64 and learning rate of 0.001). Fine-tuning was performed by initializing the network with the weights of the pre-trained model, and then all layers were fine-tuned. Moreover, the models were trained using 90% of the dataset and tested using the remaining subset. For a thorough evaluation of our solution, we conducted several experiments by varying two main parameters: a) minimum confidence threshold (C), and b) γ. The default values of C and γ were 0.5 and 1.3x, respectively. To evaluate our approaches, we report mAP (i.e., mean average precision) scores of our results.

Baseline Approach. We implemented a baseline approach adapted from the state-of-the-art material recognition method [13]. For this approach, we fine-tuned the GoogLeNet model on the image segments of □ (i.e., trained on our 18 classes of materials) that display “plain” materials (i.e., no graffiti). To obtain image parts of “plain”
In this paper, we introduced the problem of covered material recognition using a real graffiti image dataset obtained from LASAN. We first proposed straightforward approaches that utilize state-of-the-art deep learning algorithms. Thereafter, we enhanced these approaches by applying heuristics based on proportional or semantic expansions. Our experimental results demonstrated that the proposed approaches with heuristics obtained better mAP scores achieving an improvement factor of 2.1x compared with the baseline. As part of future work, we plan to integrate this solution in our visionary framework (Translational Visual Data Platform, TVDP [35]) for smart cities.

Acknowledgment

This work was supported in part by NSF grants IIS-1320149 and CNS-1461963, the USC Integrated Media Systems Center, and unrestricted cash gifts from Oracle and Google.

5.2. Evaluation Results

Baseline vs. Proposed Approaches. Fig. 4a shows the mAP scores of the baseline versus proposed approaches. In general, it is evident that an approach for plain material recognition (i.e., baseline approach) was not suitable for recognizing covered materials. In particular, the straightforward approaches (i.e., either OLA or TLA) were able to achieve an improvement factor of 0.75x compared with the baseline. Both OLA and TLA showed similar results with a negligible difference; hence OLA is preferred because it requires less training and prediction time. Moreover, OLA and TLA can be further improved by utilizing two heuristics (i.e., proportional or semantic expansion described in Section 4.2). In particular, both OLA and TLA with PE and SE were able to improve the mAP score by a factor of 1.6x and 1.8x, respectively. The experiments based on SE showed a better mAP score compared to the one based on PE because we used the annotations surrounding the entire objects marked with graffiti to obtain effective semantic expansion.

Impact of Varying γ. Fig. 4b shows the mAP scores of OLA − PE while varying γ and fixing C at 0.5. In general, PE with a small value of γ enables generating a better learned model for covered material recognition because the enlarged regions can properly depict additional visual cues of the underlying materials. However, as the value of γ grows, the model may gradually introduce more noise by learning the visual cues of other materials in addition to the underlying material of interest; hence degrading the effectiveness of the learned model. In our case study, mAP scores of the models of OLA − PE were increased using different values of γ up to 1.5.

Impact of Varying C. Fig. 4c shows the mAP scores of OLA − PE while varying C and fixing γ at 1.3x. In general, decreasing C enables the learned model to report more predictions; hence increasing the chances of reporting a correct prediction. In our study, OLA − PE showed an increasing trend of mAP when decreasing C (i.e., it obtained an mAP of 60% and 49% when using C = 0.01 and 0.6, respectively).

6. CONCLUSION

Table 1: Graffiti Dataset Distribution among the Labels of Covered Materials

<table>
<thead>
<tr>
<th>Label</th>
<th># images</th>
<th>Label</th>
<th># images</th>
<th>Label</th>
<th># images</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stone Wall</td>
<td>11,153</td>
<td>Wood Wall</td>
<td>167</td>
<td>Metal Wall</td>
<td>122</td>
</tr>
<tr>
<td>Fabric Fence</td>
<td>284</td>
<td>Stone Pole</td>
<td>227</td>
<td>WoodPole</td>
<td>186</td>
</tr>
<tr>
<td>Wood Door</td>
<td>350</td>
<td>Metal Door</td>
<td>580</td>
<td>Metal Pole</td>
<td>989</td>
</tr>
<tr>
<td>Stone Stairs</td>
<td>116</td>
<td>Wood Tree</td>
<td>275</td>
<td>Metal Box</td>
<td>735</td>
</tr>
<tr>
<td>Stone Road</td>
<td>3,409</td>
<td>Metal Fire Hydrant</td>
<td>102</td>
<td>Wood Wall</td>
<td>265</td>
</tr>
<tr>
<td>Stone Pole</td>
<td>265</td>
<td>Fabric Fence</td>
<td>270</td>
<td>WoodPole</td>
<td>270</td>
</tr>
<tr>
<td>Wood Door</td>
<td>350</td>
<td>Stone Pole</td>
<td>227</td>
<td>WoodPole</td>
<td>186</td>
</tr>
<tr>
<td>Stone Stairs</td>
<td>116</td>
<td>Wood Tree</td>
<td>275</td>
<td>Metal Box</td>
<td>735</td>
</tr>
<tr>
<td>Stone Road</td>
<td>3,409</td>
<td>Metal Fire Hydrant</td>
<td>102</td>
<td>Wood Wall</td>
<td>265</td>
</tr>
<tr>
<td>Stone Pole</td>
<td>265</td>
<td>Fabric Fence</td>
<td>270</td>
<td>WoodPole</td>
<td>186</td>
</tr>
<tr>
<td>Wood Door</td>
<td>350</td>
<td>Stone Pole</td>
<td>227</td>
<td>WoodPole</td>
<td>186</td>
</tr>
<tr>
<td>Stone Stairs</td>
<td>116</td>
<td>Wood Tree</td>
<td>275</td>
<td>Metal Box</td>
<td>735</td>
</tr>
<tr>
<td>Stone Road</td>
<td>3,409</td>
<td>Metal Fire Hydrant</td>
<td>102</td>
<td>Wood Wall</td>
<td>265</td>
</tr>
<tr>
<td>Stone Pole</td>
<td>265</td>
<td>Fabric Fence</td>
<td>270</td>
<td>WoodPole</td>
<td>186</td>
</tr>
<tr>
<td>Wood Door</td>
<td>350</td>
<td>Stone Pole</td>
<td>227</td>
<td>WoodPole</td>
<td>186</td>
</tr>
<tr>
<td>Stone Stairs</td>
<td>116</td>
<td>Wood Tree</td>
<td>275</td>
<td>Metal Box</td>
<td>735</td>
</tr>
<tr>
<td>Stone Road</td>
<td>3,409</td>
<td>Metal Fire Hydrant</td>
<td>102</td>
<td>Wood Wall</td>
<td>265</td>
</tr>
</tbody>
</table>

(a) Baseline vs. Proposed Approaches

(b) OLA − PE w/ Varying γ

(c) OLA − PE w/ Varying C

Fig. 4: Proposed Approaches Evaluation

Fig. 5: Annotation Example

Impact of Varying γ. Hereafter, we report the results only for OLA − PE as the same trends hold for other approaches.

---

4OLA obtained on average 16% and 28% improvement percentages w.r.t. TLA in terms of training and prediction times, respectively. The detailed performance comparison is omitted due to the space limitation.

5The effectiveness of SE depends on obtaining accurate segmentation when using a segmentation algorithm (e.g., [32]).
References


