Material Recognition in the Wild with the Materials in Context Database

Sean Bell*, Paul Upchurch*, Noah Snavely, Kavita Bala
Department of Computer Science, Cornell University

Recognizing materials in real-world images is a challenging task. Real-world materials have rich surface texture, geometry, lighting conditions, and clutter, which combine to make the problem particularly difficult. In this paper, we introduce a new, large-scale, open dataset of materials in the wild, the **Materials in Context Database** (MINC), and combine this dataset with deep learning to achieve material recognition and segmentation of images in the wild. MINC is available online at http://minc.cs.cornell.edu/.

MINC is an order of magnitude larger than previous material databases, while being more diverse and well-sampled across its 23 categories. Using MINC, we train convolutional neural networks (CNNs) for two tasks: classifying materials from patches, and simultaneous material recognition and segmentation in full images. For patch-based classification on MINC we found that the best performing CNN architectures can achieve 85.2% mean class accuracy. We convert these trained CNN classifiers into an efficient fully convolutional framework combined with a fully connected conditional random field (CRF) to predict the material at every pixel in an image, achieving 73.1% mean class accuracy. Our experiments demonstrate that having a large, well-sampled dataset such as MINC is crucial for real-world material recognition and segmentation.

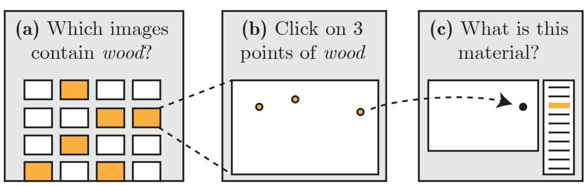

Figure 1: **AMT pipeline. (a)** We filter out images that contain a certain material, **(b)** workers click on materials, and **(c)** workers validate click locations by re-labeling each point. Example responses are shown in orange. Using this three-stage pipeline we collected over 2.3 million material labels from Flickr and Houzz images.

Architecture	Validation	Test
AlexNet	82.2%	81.4%
GoogLeNet	85.9%	85.2%
VGG-16	85.6%	84.8%

Table 1: **Patch material classification results.** Mean class accuracy for different CNNs trained on 2.5 million patches from our new dataset, MINC.

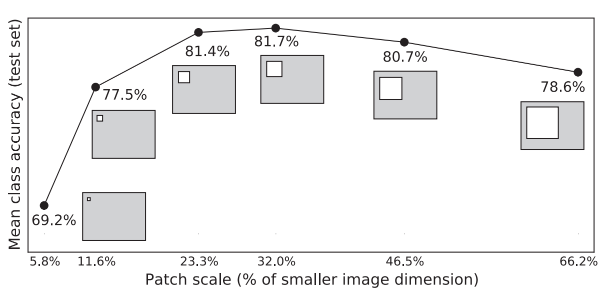

Figure 2: **Optimal patch size on MINC**. When training a CNN to predict materials, the optimum subregion around each labeled point is a trade-off between context and spatial resolution.

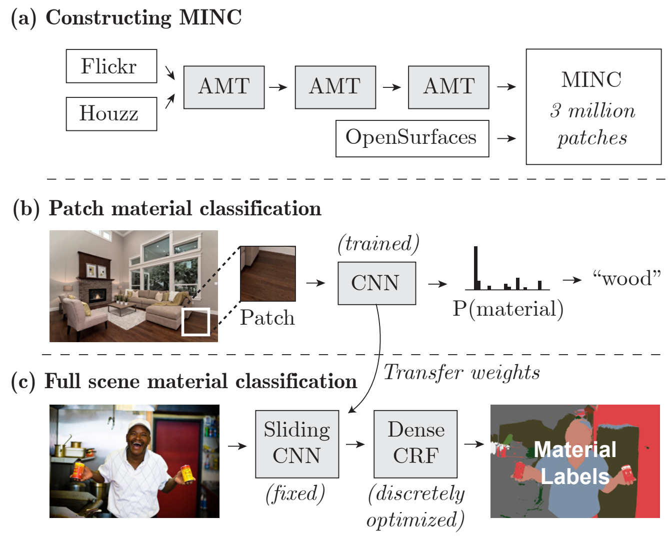

Figure 3: **Overview. (a)** We construct a new dataset by combining OpenSurfaces [1] with a novel three-stage Amazon Mechanical Turk (AMT) pipeline. **(b)** We train various CNNs on patches from MINC to predict material labels. **(c)** We transfer the weights to a fully convolutional CNN to efficiently generate a probability map across the image; we then use a fully connected CRF to predict the material at every pixel.

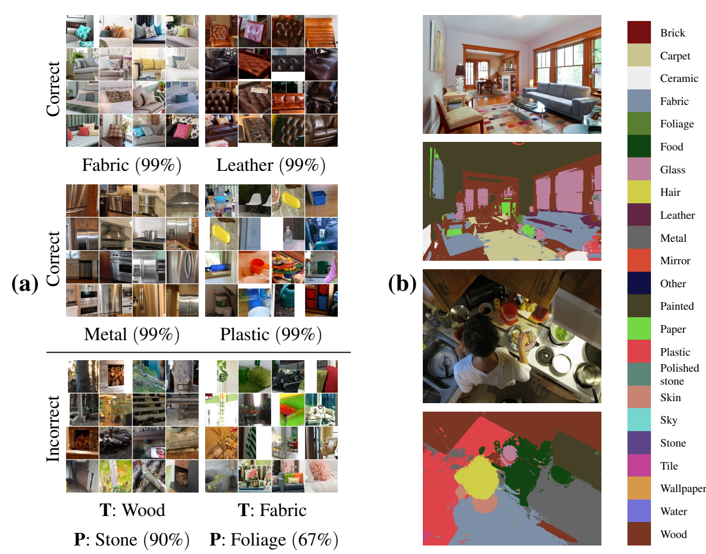

Figure 4: **Results. (a)**: high confidence single patch predictions. Top rows: correct predictions. Bottom row: incorrect predictions (T: true, P: predicted). Percentages indicate confidence (the predictions shown are at least this confident). **(b)**: full-scene material classifications with GoogLeNet (average pooling layer removed).

[1] Sean Bell, Paul Upchurch, Noah Snavely, and Kavita Bala. OpenSurfaces: A richly annotated catalog of surface appearance. *ACM Trans. on Graphics (SIGGRAPH)*, 32(4), 2013.

* Authors contributed equally
This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.