BWMP2: DATASET RGB PARA CLASIFICACIÓN DE MATERIALES CON UN MODELO FUNDACIONAL FINAMENTE AJUSTADO

Juan José Calderón Gómez
Ingeniería de sistemas, Universidad Industrial de Santander, Colombia,
juan2220049@correo.uis.edu.co

Brayan Sneider Sánchez Muñoz
Ingeniería de sistemas, Universidad Industrial de Santander, Colombia,
brayan2220083@correo.uis.edu.co

César Darío Vanegas Oviedo
Ingeniería de sistemas, Universidad Industrial de Santander, Colombia,
cesar2220040@correo.uis.edu.co

Dana Meliza Villamizar Lizarazo
Ingeniería de sistemas, Universidad Industrial de Santander, Colombia,
dana2220081@correo.uis.edu.co

Nelson Fabián Pérez Pérez
Ingeniería de sistemas, Universidad Industrial de Santander, Colombia,
nelson2200183@correo.uis.edu.co

Hoover Fabián Rueda Chacón, PhD.
Ingeniería de Sistemas, Universidad Industrial de Santander, Colombia,
hfarueda@uis.edu.co

Objetivo: Crear y evaluar el conjunto de datos *BWMP2* para clasificar materiales (ladrillo, madera, metal, papel y plástico) usando un modelo fundacional finamente ajustado. **Metodología:** Se crea una base de datos de imágenes propias tomadas con un dispositivo móvil, posteriormente se hace la estratificación del conjunto y se utiliza ResNet-50 preentrenada ajustada al conjunto de datos creado: *BWMP2* (150 imágenes RGB). Se adaptaron capas lineales y se mantuvieron congeladas las capas convolucionales. El modelo se cuantizó para web usando *transformers.js*. **Resultados:** El modelo finamente ajustado alcanzó una precisión media del 93.3% en la clasificación de los materiales seleccionados, el modelo final se desplegó a plataformas web con un tamaño de 24.9 MB tras su cuantización. **Conclusiones:** Se ofrece un conjunto de datos público y un modelo eficaz y liviano con alta precisión abriendo la puerta a futuras mejoras del modelo y expansión del conjunto de datos.

Palabras clave: Clasificación de materiales, Ajuste fino, Modelos fundacionales, Aprendizaje profundo, MLOps.

1. Introducción

La clasificación de materiales (*Material Classification*) es una tarea fundamental dentro del campo de la visión por computadora (Sticlaru, 2017), que consiste en identificar y categorizar diferentes tipos de materiales a partir de imágenes, asignándoles una etiqueta específica que represente su composición, por ejemplo, ladrillo, madera, metal, papel o plástico (Upchurch & Niu, 2022).

Esta capacidad es crucial para una amplia variedad de aplicaciones industriales, como el reciclaje (Saponaro et al., 2015) o la robótica (Bednarek et al., 2019), ya que permite a las máquinas reconocer y diferenciar materiales de forma automática.

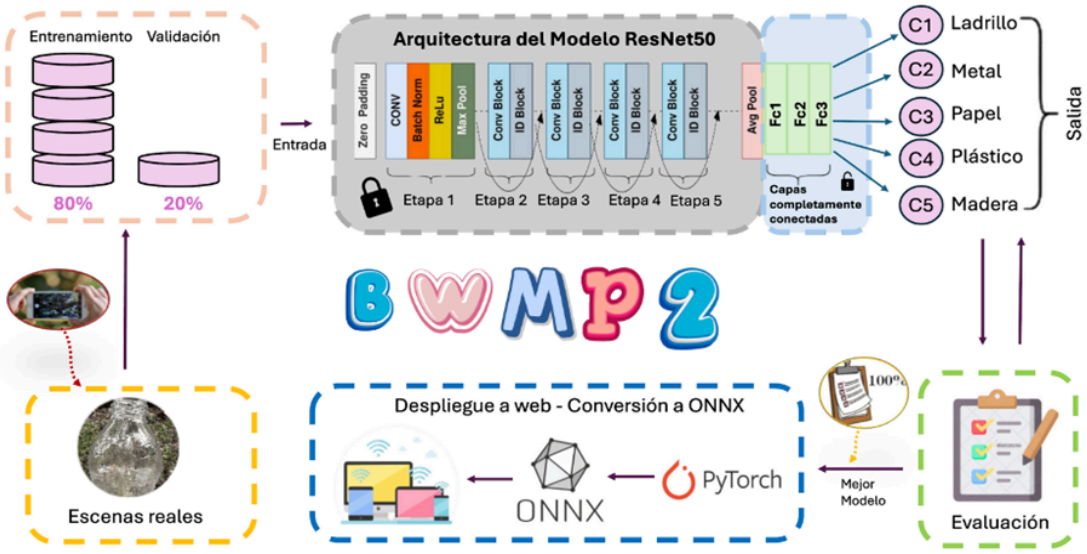
Una de las principales aplicaciones de la clasificación de materiales es en la gestión automatizada de residuos (Saponaro et al., 2015). En las plantas de reciclaje, la identificación y separación de materiales como plástico, metal o papel son procesos muy útiles a la hora de optimizar el reciclaje y reducir los desechos no reciclables (Chen, 2021).

Sistemas automáticos de clasificación permiten incrementar la eficiencia de las plantas en la separación de materiales (Chen et al., 2021). Otra aplicación relevante es la robótica industrial, donde la clasificación de materiales puede ayudar a los robots a tomar decisiones informadas sobre cómo manipular objetos (Bednarek et al., 2019).

Los enfoques utilizados previamente en la clasificación de materiales abarcan diversas metodologías basadas en características físicas y ópticas. Las imágenes térmicas se emplean para detectar variaciones de temperatura en los materiales, útiles en la robótica y reciclaje (Kerr et al., 2013). Las imágenes de profundidad y el LIDAR capturan información tridimensional, optimizando la identificación de formas y texturas (Han et al., 2023).

Los métodos espectrales analizan las firmas de los materiales mediante longitudes de onda específicas (Ibrahim, 2010), mientras que la técnica BRDF (Bidirectional Reflectance Distribution Function) evalúa cómo los materiales reflejan la luz en distintas direcciones, utilizada en modelado de superficies y simulaciones (Liu & Gu, 2014).

Figura 1. Arquitectura del modelo presentado



Fuente: Elaboración propia

En nuestro trabajo, creamos un *dataset* RGB específico para esta tarea, capturando imágenes de cinco materiales: ladrillo, madera, metal, papel y plástico. Utilizamos la arquitectura del modelo fundacional ResNet-50 (He et al., 2016), preentrenada en ImageNet, y la adaptamos mediante la técnica de ajuste fino para mejorar su capacidad de reconocer estos materiales. Posteriormente, desplegamos el modelo en entornos web, convirtiéndolo al formato ONNX (*Open Neural Network Exchange*), lo que permite su uso en diferentes plataformas.

2. Metodología

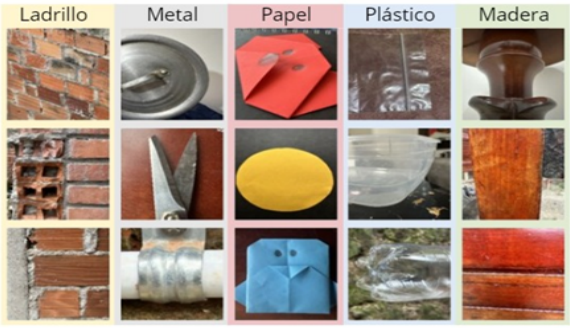
Para abordar la clasificación de materiales utilizando imágenes RGB, usamos la arquitectura del modelo fundacional ResNet-50, preentrenada en ImageNet (Yosinski et al., 2014), que se adaptó mediante la técnica de *fine-tuning*. En la Figura 1 se ilustra la estructura del método.

2.1 Creación del Dataset

La base de este trabajo es la creación de un conjunto de datos RGB específico para la tarea de clasificación de materiales, que hemos llamado **BWMP2** *(brick, wood, metal, paper, plastic),* dado que los datasets disponibles para esta tarea tienen limitaciones significativas en términos de diversidad y volumen de imágenes. Por esto, capturamos nuestro propio dataset utilizando un dispositivo móvil de uso general, específicamente, el iPhone 14 Pro, el cual cuenta con un sensor de 48MP y una apertura de f/1.78 (Apple Inc., 2022).

Capturamos un total de 150 imágenes de cinco materiales representativos: ladrillo, madera, metal, papel y plástico, cada material cuenta con 30 imágenes seleccionadas por su presencia en el entorno diario y su importancia en aplicaciones industriales. En la Figura 2 se ilustran algunas muestras del *dataset*.

Figura 2. Imágenes extraídas de BWMP2



Fuente: Elaboración propia

Las imágenes adquiridas se obtuvieron en el formato HEIC (Hannuksela et al., 2015), se convirtieron al formato PNG y allí fueron redimensionadas a 256x256, un formato comúnmente utilizado en tareas de visión por computadora que balancea la eficiencia computacional y la preservación de detalles visuales. De estas 150 imágenes, 120 fueron destinadas al conjunto de entrenamiento y 30 al conjunto de prueba, asegurando una estratificación uniforme, de modo que cada clase de material estará representando el 20% de los datos en ambos subconjuntos. Una vez completada la fase de captura y procesamiento de imágenes, el dataset se subió a la plataforma *Hugging Face*. El dataset se encuentra públicamente disponible en (Hands-On Computer Vision, 2024).

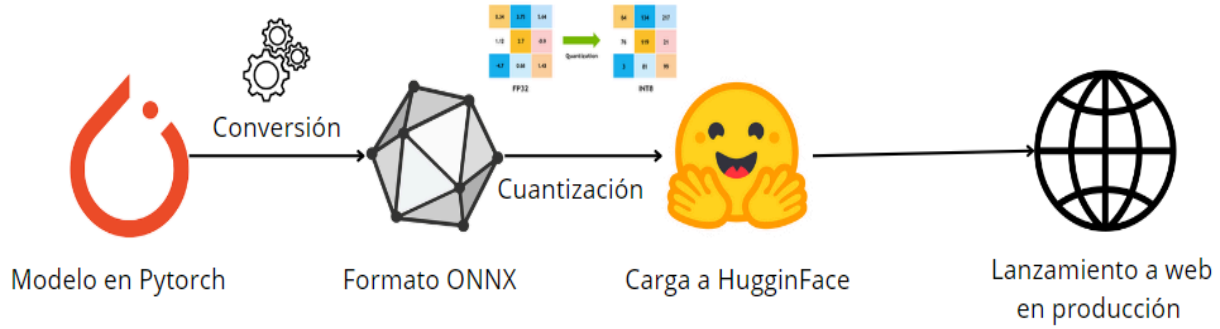
2.2 Modelo y Ajuste fino

Se empleó un modelo de aprendizaje profundo, específicamente ResNet-50, una arquitectura que ha mostrado resultados sobresalientes en tareas de clasificación de imágenes. ResNet-50 ha sido preentrenada con el conjunto de datos ImageNet (Yosinski et al., 2014), lo que proporciona un punto de partida sólido dado que los pesos ya han aprendido representaciones de características visuales fundamentales.

El procedimiento de ajuste fino (*fine tuning*) se basa en la calibración de parámetros seleccionados que generan cambios (ya sea de desempeño, rendimiento o funcionalidad) sobre la arquitectura estándar de ResNet-50 (The University of Groningen & Friederich, 2017). En el modelo presentado, este proceso consistió en modificar la arquitectura eliminando su capa lineal final, reemplazándola con tres capas completamente conectadas de 512, 256 y 5 neuronas, respectivamente.

Esta estructura en forma de embudo sigue un diseño de *bottleneck* que consiste en la reducción y restauración de dimensiones.

Figura 3. Proceso del despliegue del modelo a producción.



Fuente: Elaboración propia

Es usado en las capas finales de las arquitecturas convolucionales, donde se modifican las dimensiones para solucionar problemas de complejidad computacional y tiempo de procesamiento debido a su profundidad (He et al., 2016).

Las capas convolucionales de ResNet-50 se mantuvieron congeladas, es decir, sus parámetros no se actualizaron durante el entrenamiento, dejando aproximadamente 1.1 millones de parámetros entrenables.

Durante el entrenamiento, las imágenes de entrada se normalizaron usando la media y desviación estándar de ImageNet, una práctica común cuando se ajustan modelos preentrenados (Yosinski et al., 2014). Se definieron valores de media y desviación estándar comúnmente utilizados al momento de trabajar con ImageNet, y estos se aplican sobre RGB.

2.3 Cuantización para Inferencia en Web

Tras el entrenamiento del modelo, se optimizó para que fuera más eficiente en términos de espacio y velocidad para su despliegue en plataformas web. La conversión del modelo entrenado a formato ONNX (*Open Neural Network Exchange*), el cual es un formato de código abierto para representar modelos de inteligencia artificial (ONNX, n.d.), logró optimizarlo para inferencia en diferentes entornos, especialmente en dispositivos con recursos limitados como navegadores web.

El modelo en formato *float16* ofrece un buen balance entre precisión y peso, pero para mejorar aún más su rendimiento en la web procedimos a la cuantización del modelo a *uint8*, para lo cual utilizamos *ONNX Runtime*, que es un acelerador de modelos de machine learning multiplataforma (*ONNX Runtime*, 2023).

Con este, se realizó un proceso de cuantización dinámica, técnica utilizada después de que el modelo ha sido entrenado, para reducir su tamaño y acelerar las inferencias (Z. Liu et al., 2022).

Este proceso inicialmente recibe el modelo preentrenado y transforma el valor de los pesos, pero no el de las activaciones, ya que estas varían durante la inferencia dependiendo de los datos de entrada, por lo que su calibración resulta desafiante.

Para manejar esto se incluyen funciones que observan los rangos de activaciones en tiempo real (durante la inferencia) y las cuantizan, optimizando las operaciones matemáticas y ganando precisión, al costo de una eficiencia reducida. Este proceso de cuantización reduce significativamente el tamaño del modelo, lo cual es crucial para aplicaciones web donde el tiempo de descarga y la capacidad de procesamiento son factores limitantes.

La cuantización a *uint8* implica una reducción en la capacidad de representación del modelo, lo que potencialmente puede afectar la precisión de la clasificación. Sin embargo, estudios previos han demostrado que los modelos de clasificación de imágenes son particularmente robustos a la cuantización sin que se produzcan caídas significativas en el rendimiento, como describen Rokh et al. (2023, p. 41).

2.4 Despliegue del Modelo

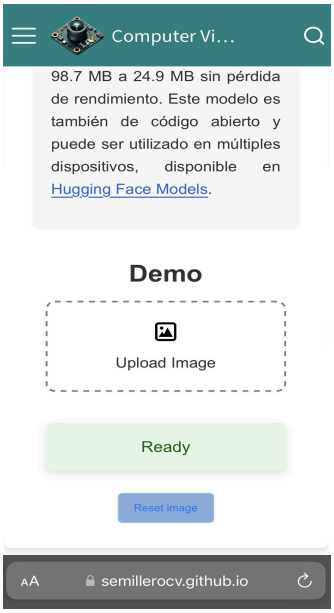
El despliegue del modelo en plataformas web se realizó utilizando la librería *transformers.js,* una herramienta poderosa y versátil que permite la ejecución de modelos ONNX directamente en el navegador del usuario, eliminando la necesidad de un servidor con back-end.

Desarrollamos una página web específica para este propósito, públicamente disponible en la página del semillero *Hands-On Computer Vision* (Hands-On Computer Vision, 2024b), a través de ella los usuarios pueden interactuar con el modelo en tiempo real, cargando imágenes desde su dispositivo o activando la cámara para realizar inferencias en directo.

Para lograr esto, se implementó un código en JavaScript que no solo facilita la carga del modelo, sino que gracias al uso de un modelo cuantizado optimiza el tiempo de respuesta, lo que mejora la experiencia del usuario incluso en plataformas con recursos limitados, como dispositivos móviles.

En la Figura 4 se presenta una captura de la página donde se aprecia el diseño interactivo y la opción de carga de imágenes.

Figura 4. Captura de la página web para el despliegue del modelo en producción.



Fuente: Elaboración propia

3. Resultados

Se presentan los resultados obtenidos tras el entrenamiento y evaluación del modelo finamente ajustado ResNet - BWMP2 para la clasificación de materiales. Los resultados se analizan en términos de *accuracy* por clase y *mean accuracy* sobre todo el conjunto de prueba.

La *accuracy* se define como la proporción de predicciones correctas sobre el total de predicciones realizadas. Formalmente, el accuracy para una determinada clase se calcula como:

$$accuracy = \frac{TP}{TP + FN} \qquad (2)$$

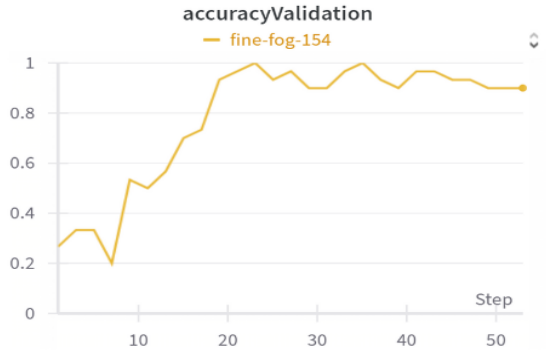
donde, *TP* (verdaderos positivos) son el número de predicciones correctamente clasificadas como la clase de interés y *FN* (falsos negativos) instancias de la clase que fueron clasificadas como otra clase.

La *mean accuracy* (Macc) representa el promedio de las precisiones obtenidas por cada clase y se calcula de la siguiente manera:

$$mean\ accuracy = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FN_i} \qquad (3)$$

donde N es el número total de clases (en nuestro caso, 5, dado por : ladrillo, madera, metal, papel, y plástico).

Figura 5. *Mean accuracy* para validación.



Fuente. Datos visualizados utilizando *Weights & Biases* (WandB).

En la Figura 5 se observa que la *mean accuracy* del modelo tiende a acercarse a 1 conforme aumentan las épocas. Esto indica que el modelo está mejorando su capacidad para hacer predicciones correctas, logrando así un rendimiento más preciso y consistente. En este trabajo, se compararon tres variantes de la red ResNet-50: *sin ajuste fino*, *ajuste fino completo*, y *ajuste*

fino con capas convolucionales congeladas. La Tabla 1 muestra los resultados.

Tabla 1. Resultados para las tres variantes de la red ResNet-50.

Modelo	Macc	Parámetros
Sin ajuste fino	0.2333	24.689.733
Ajuste fino completo	0.5333	24.689.733
Ajuste fino con capas convolucionales congeladas	**0.9333**	**1.181.701**

Como se observa en la Tabla 1, el modelo "ResNet-50 sin ajuste fino" tiene un rendimiento significativamente más bajo en comparación con las versiones ajustadas. Esto es porque los pesos de la red preentrenada están optimizados para la clasificación de objetos generales, no para la clasificación de materiales.

Por otro lado, el modelo "ResNet-50 con ajuste fino completo" alcanza una *mean accuracy* del 53.3%. Esto se debe a que todas las capas de la red fueron actualizadas durante el entrenamiento, permitiendo que el modelo ajustara sus parámetros para aprender características específicas de cada material.

Finalmente, el modelo "ResNet-50 con capas convolucionales congeladas" muestra un rendimiento significativamente mejor que la versión sin ajuste fino y la de ajuste fino completo. Este modelo logró una *mean accuracy* del 93.3%, lo que demuestra que incluso con las capas convolucionales congeladas, el ajuste fino de las capas lineales permite obtener una mejora. En la Tabla 2 se muestran los resultados por clase del modelo finamente ajustado y congelado.

Tabla 2. Accuracy por clase para el modelo finamente ajustado y congelado.

Ladrillo	Madera	Metal	Papel	Plástico
0.998	0.667	0.885	0.833	1

Las diferencias en los resultados, como se observa en la Tabla 2, pueden explicarse en gran parte por las características del conjunto de imágenes utilizado. Las clases *plástico* y *ladrillo* muestran una alta precisión, que podría ser atribuida a que sus imágenes presentan características visuales distintivas y fácilmente diferenciables, como colores y texturas.

Se evaluaron los siguientes parámetros de complejidad computacional del modelo: FLOPs totales (21.354 GigaFLOPs), iteraciones por segundo (67.47), tiempo de convergencia (43 s) y máximo de memoria utilizada (327.41 MB). Estos indicadores reflejan la eficiencia y recursos requeridos durante el entrenamiento. Para el

entrenamiento se utilizó un tamaño de lote de 64 imágenes, el modelo fue entrenado durante 30 épocas, usando una GPU NVIDIA T4. Finalmente, mediante el proceso de cuantización se logró reducir el peso del modelo significativamente de 98.7 MB a 24.9 MB (modelo cuantizado).

4. Discusión

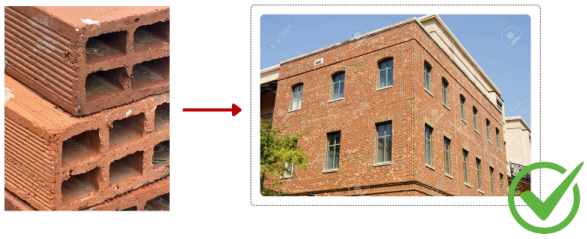
Analizando el desempeño del modelo ResNet-BWMP2, se observa una alta precisión en las clases de ladrillo y plástico, lo cual puede atribuirse a la presencia de características visuales distintivas en estas clases. En contraste, las clases de madera, papel y metal mostraron una menor precisión, lo que puede deberse a la mayor variabilidad en sus características visuales presentes durante la toma del dataset.

El modelo demuestra capacidad para reconocer imágenes en contextos reales. Se realizó esta prueba dado que el conjunto de datos incluye únicamente imágenes del objeto de interés, sin contextos o detalles adicionales. Como se ilustra en la Figura 6, el modelo ha sido capaz de identificar correctamente un edificio como ladrillo, evidenciando su habilidad para generalizar.

5. Conclusiones

Este trabajo ofrece un conjunto de datos y un modelo ligero en código abierto para clasificación de materiales, fomentando el avance y la colaboración en visión por computadora. Futuras investigaciones abordarán las limitaciones actuales y ampliarán el enfoque hacia segmentación y detección de objetos, mejorando las capacidades del modelo y su impacto.

Figura 6. Prueba con escenas reales.



Fuente. Elaboración propia

6. Agradecimientos

Agradecemos al Semillero de Investigación Hands-On Computer Vision por su valiosa dirección y apoyo durante el desarrollo de este proyecto.

Referencias

Bednarek, J., Bednarek, M., Kicki, P., & Walas, K. (2019). Robotic touch: Classification of materials for manipulation and walking. *2019 2nd IEEE International Conference on Soft Robotics (RoboSoft)* http://dx.doi.org/10.1109/robosoft.2019.8722819

Chen, H. (2021). Optimization of an intelligent sorting and recycling system for solid waste based on image recognition technology. *Advances in Mathematical Physics*,

2021, 1–12. https://doi.org/10.1155/2021/4094684

Han, Y., Salido-Monzú, D., & Wieser, A. (2023). Classification of material and surface roughness using polarimetric multispectral LiDAR. *Optical Engineering*, *62*(11). https://doi.org/10.1117/1.oe.62.11.114104

Hands-On Computer Vision. (2024). *CV*. Semillero Hands-on Computer Vision. https://semillerocv.github.io/proyectos/proyecto2.html

He, K., Zhang, X., Ren, S., & Sun, J. (2016, June). Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. http://dx.doi.org/10.1109/cvpr.2016.90

Ibrahim, A. (2010). Spectral imaging method for material classification and inspection of printed circuit boards. *Optical Engineering*, *49*(5), 057201. https://doi.org/10.1117/1.3430606

Kerr, E., McGinnity, T. M., & Coleman, S. (2013). Material classification based on thermal properties — A robot and human evaluation. *2013 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, http://dx.doi.org/10.1109/robio.2013.6739602

Liu, C., & Gu, J. (2014). Discriminative illumination: Per-Pixel classification of raw materials based on optimal projections of spectral BRDF. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *36*(1), https://doi.org/10.1109/tpami.2013.110

(n.d.). ONNX | Home. https://onnx.ai/index.html

ONNX runtime. (2023). Onnxruntime. https://onnxruntime.ai/docs/

sen, S., Kolagunda, A., & Kambhamettu, C. (2015). Material classification with thermal imagery. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4649–4656. http://dx.doi.org/10.1109/cvpr.2015.7299096

Sticlaru, A. (2017, October 17). *Material Classification using Neural Networks*. arXiv.Org. https://arxiv.org/abs/1710.06854

Upchurch, P., & Niu, R. (2022). A Dense Material Segmentation Dataset for Indoor and Outdoor Scene Parsing. In *Lecture Notes in Computer Science* (pp. 450–466). Springer Nature Switzerland. http://dx.doi.org/10.1007/978-3-031-20074-8_26