2025 European Conference on Computing in Construction

CIB W78 Conference on IT in Construction

Porto, Portugal
July 14-17, 2025

MATERIAL DETECTION AND CLASSIFICATION IN 2D ARCHITECTURAL DRAWINGS USING COMPUTER VISION FOR MATERIAL PASSPORTS

Ainur Kairlapova[1], Kasimir Forth[1,2], Andrea Carrara[1], and André Borrmann[1]
[1]Technical University of Munich, Munich, Germany
[2]ETH Zürich, Zürich, Switzerland

Abstract

Retrofitting Europe's ageing buildings is critical for sustainable development, yet most existing structures lack comprehensive building models. Architectural drawings often include material information, but the level of detail is insufficient for accurately assessing retrofit potential. Manual extraction of this data is costly and inefficient, creating challenges for digitalizing architectural data. This paper proposes a novel Computer Vision (CV) approach to automate material detection and classification in drawings for simplified Material Passports (MP). Three CNN architectures—U-Net, ResNet50, and MobileNetV2—are evaluated, with MobileNetV2 performing best on mixed-source datasets. Case studies highlight its potential in facilitating resource-efficient retrofits, particularly for reinforced concrete.

Introduction

The ageing building infrastructure presents a significant challenge, as a substantial share of existing structures now require renovation and refurbishment. These activities are not limited to mere "re-building" but must also prioritize minimizing the footprint of buildings (European Commission, 2020). According to a technical report by the Joint Research Centre, construction and demolition waste is responsible for almost 40% of all waste generated in the EU (García et al., 2023). This context emphasizes the importance of key concepts in the construction sector to facilitate the transition to a Circular Economy (CE). Circularity focuses on extending the lifecycle of materials, Life Cycle Assessment (LCA) evaluates the environmental impact of products from creation to disposal, and Material Passports (MP) store data about materials to enable reuse and recycling.

Current approaches for creating inventories or MPs of existing buildings either focus on capturing and automatically generating 3D models of visible components and materials, also known as Scan-to-BIM. Or they use manual workflows (Honic et al., 2020), which is costly and time-intensive. However, hidden primary structural materials are rarely included in these approaches. Furthermore, to the best of the authors' knowledge, automating technical drawings is not used yet to create MPs of existing buildings, while these data sources represent domain-specific information using geometric and semantic material information using hatchings.

The primary objective of this paper is to develop a methodology for the detection of primary structural materials from two-dimensional (2D) architectural drawings. The proposed methodology aims to create a simplified MP that encapsulates essential information about the geometry and type of materials used in building floor plans and sections.

Background and related works

In this section, we first introduce general industry practices on material representations. Next, we explore widely recognized architectures of Convolutional Neural Networks (CNN) for Computer Vision (CV) tasks. Finally, we present current applications of using CV for architectural drawings.

Material representation

Technical drawings serve as visualization, communication or documentation in various engineering disciplines. Common graphical representation allows smooth communication between experts in the architectural and civil engineering fields. A common way to represent the material is using relevant hatching. One of the technical drawing standards is ISO 128-3:2022, which describes in detail the technical representation of products (ISO 128-3:2022, 2022). The standard highlights that the hatching should be distinguished from its principal outlines. In the case of a rectangular shape, the hatch angle should be 45 degrees (ISO 128-3:2022, 2022). For the specific materials, the representation shall be indicated on the drawing itself either as a legend or using a reference to another document.

The work of Ernst Neufert's "Bauentwurfslehre" is considered one of the comprehensive compilations of the building design principles (Neufert et al., 2019). Regarding hatching, the book emphasizes its importance in communicating material properties, construction details, and spatial separations within architectural drawings. It provides systematic guidance on applying hatching to standardize visual representations, ensuring clarity and uniformity in plans.

Convolutional neural networks for computer vision

The underlying technology of deep learning (DL) is neural networks. To compute the output of one layer, the input matrix is multiplied by the weight of that layer, and after that, the activation formula is applied to produce the output. Hidden layers between the input and output layers raise the level of complexity of tasks that can be solved. The applied layers and activation functions can "learn" the non-linear relations in the data and may increase the complexity of solvable problems. The loss function is defined to reach the optimum output provided by training data, which reflects an error at the output layer (Russell and Norvig, 2021). Weights are adapted to minimise the loss function after each epoch (iteration) through the data set.

MobileNetV2 is one of the neural network architectures that is characterized by its accuracy and efficiency in mobile applications (Sandler et al., 2019). The U-Net architecture was introduced specifically for image segmentation tasks to reach precise results with a small data set (Ronneberger et al., 2015). ResNet is a neural network architecture designed to learn residual functions with respect to layer inputs(He et al., 2015).

CV in architectural drawings

With the growth of computing capacity, the possibilities of ML are expanding into various fields. In the Architecture, Engineering, Construction (AEC) industry, the research field of artificial intelligence (AI) emerged in 1956 (Darko et al., 2020). Since then, optimization techniques, genetic algorithms, neural networks and other related topics have garnered significant attention. DL methods and particularly CNN are widely applied in image recognition and classification as they remove the necessity for manual extraction and identification of features (Darko et al., 2020).

Mafipour et al. (2023) introduces a method to simplify the parametric modelling of bridges from technical drawings by identifying 14 key classes essential for the geometric modelling of single-span bridges. Synthetic and actual drawings are combined to train a DL model that detects elements with a mean average precision (mAP) of 89.15%, while a pre-trained model is used to extract text and dimensions (Mafipour et al., 2023).

Xiao et al. (2020) utilizes a fully convolutional neural (FCN) network for the semantic segmentation of elements in 2D drawings. The DL network achieves an 80% accuracy rate and fast operating speed, however, there is still potential for optimization by using pre-trained models and diversification of the data set (Xiao et al., 2020).

Carrara et al. (2024) introduces a novel approach to converting and reconstructing mixed-use building layouts into vectors and 3D models using two neural networks: one for semantic segmentation of walls and another for detecting openings. The method effectively handles complex layouts, incorporating features like vertical transportation spaces and non-standard forms.

Most studies concentrate on geometrical, numerical, and textual information, as well as the detection of specific elements like doors and windows. The existing approaches overlook the significance of materials, which is crucial in the context of circularity assessments and the redesign of existing building stocks.

Methodology

Overall framework

The proposed overall framework enables automated material detection and classification using architectural drawings of different data sources, such as hand-, Computer-Aided Design (CAD)-drawn, or Building Information Model (BIM)-generated. Architectural floor plans and sections are utilized as primary tools to delineate material boundaries within structural elements. These graphical outlines visually represent the primary structural materials used, which are then digitized using CV techniques. This transformation involves quantifying the graphical outlines with precision and converting them into a digital format that accurately represents the material classes of building components.

Figure 1 outlines the framework in three stages. The first stage ensures meticulous data preparation to maintain data set quality and relevance for training DL models. The second stage employs semantic segmentation using advanced CNN models, categorizing each pixel in the drawings into segments corresponding to different materials. Finally, post-processing involves fine-tuning the best-performing CNN model to enhance accuracy and reliability. The validation of the overall method is out of the scope of this paper.

Semantic segmentation

The semantic segmentation stage includes the iterative configuration of selected CNN models. Several base CNN architectures are selected to evaluate their suitability for the detection task. Key parameters, including input shape, number of classes, and other necessary configurations, are customized for optimal performance on the prepared data set.

At the initial stage, pre-trained CNN models are evaluated with frozen weights to preserve their learned features, ensuring a fair comparison tailored to the data set's characteristics. We evaluate the different CNN models both on specific data sources and a data set of mixed data sources to assess model performance on specific styles and their adaptability across diverse styles.

The annotated and pre-processed data sets are then fed into each model for training, enabling the CNNs to learn the distinctive features of materials represented in architectural drawings. All models are compiled with consistent settings, including the sparse categorical cross-entropy loss function, and uniform evaluation metrics. Adam optimizer is chosen, as it is widely used in semantic segmentation tasks, and its effectiveness in practice is shown in empirical results (Kingma and Ba, 2017). Training spans

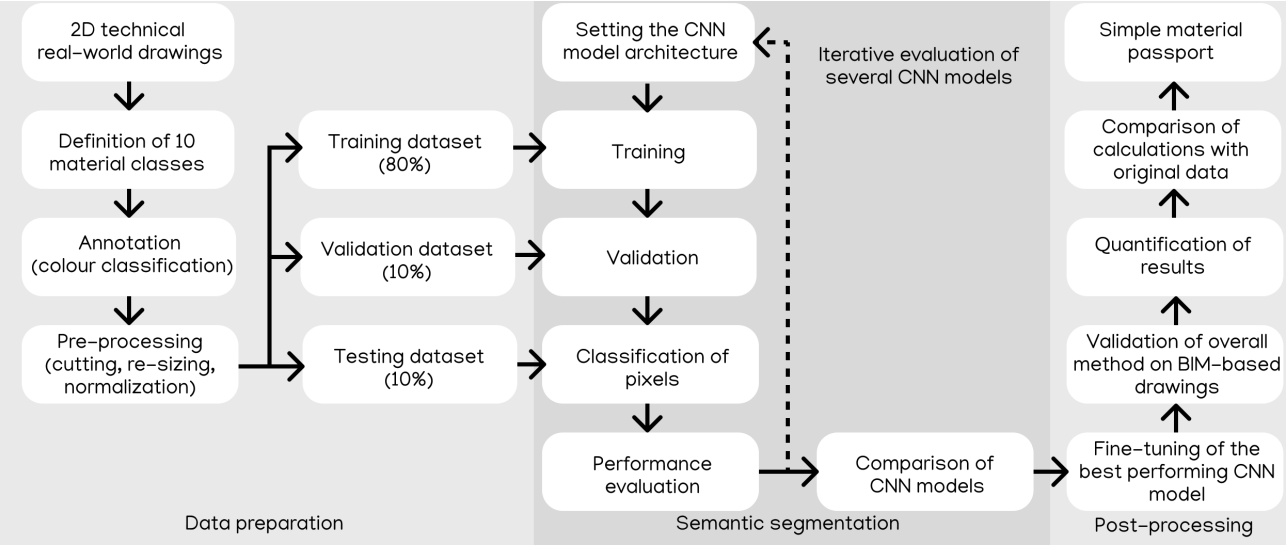

Figure 1: Proposed overall framework

a fixed number of epochs, during which models iteratively update their parameters to improve segmentation accuracy.

The model performances are evaluated using quantitative metrics such as Overall Accuracy (OA), Intersection over Union (IoU), and Mean IoU. Qualitative insights are obtained by visually inspecting predicted material masks, while comparisons with ground truth annotations serve as benchmarks for validating prediction accuracy. These steps identify the most accurate model for further refinement and result processing.

Model fine-tuning

The fine-tuning stage includes hyper-parameter optimization (batch size, number of epochs, learning rate), weight regularization, dropout and testing of class weighting and data augmentation. Hyper-parameter optimization can be automated by several open-source frameworks, such as Optuna, Ray Tune, and Keras Tuner, as the configuration of CNN model aspects can be complex and interdependent. Utilizing various loss functions can also enhance CNN model training by emphasizing different aspects of the network's performance and guiding the model to learn more effectively from diverse features.

The strategy of testing several combinations of the learning rate and batch size is used to find the most optimal. Depending on the optimizer and stability of the CNN model during training, either the linear scaling rule or the square root method needs to be used in learning rate and batch size adaptation. The class imbalance is not addressed during the initial CNN model iterations. However, class weighting is tested during the refinement phase to improve CNN model performance on underrepresented classes. This is achieved by modifying the loss function to assign higher weights to the minority classes. This adjustment encourages the CNN model to focus more on the minority class, such as different hatching during training.

Data augmentation is another critical aspect of CNN model refinement. To diversify the range of detected patterns and enhance the CNN model's adaptability, images and their corresponding masks are subjected to transformations such as rotation and zoom. This process increases the variability in the dataset, which can lead to a more robust and adaptable CNN model.

As the training data contains large amounts of irrelevant information for the task of material detection, such as backgrounds and certain architectural elements, increasing the complexity of the CNN model can lead to overfitting. It occurs when the CNN model learns the training data too well, including its noise and specific features, and consequently performs poorly on new, unseen data. Several strategies can be employed to address overfitting, including weight regularization and dropout.

Case Studies and Implementation

In this section, we first introduce the three case studies that differentiate the data sources. Next, we briefly describe the implementation of the pre-processing steps, including material class definition. Finally, the semantic segmentation and CNN fine-tuning steps are explained.

Case study selection

The dataset for this study comprises 2D architectural drawings from the floor plans of three different case studies differentiated by their data source: hand-drawn (14 drawings), CAD (12 drawings), and BIM (12 drawings). For the purposes of this study, the buildings and their data sets will be referred to as CAD, BIM, and hand-drawn, based on their assumed origins. These drawings, created between 1990 and 2010, encompass different commercial, office, and residential spaces ranging from two to five stories. We published the patches of the original drawings and the labeled ones as an open-source dataset[1].

[1] https://doi.org/10.14459/2025mp1766395

On the contrary, the drawings of a hand-drawn case study consist of hand-drawn, scanned plans. Here, the main difficulties are specific signs (for slits, chimneys, etc), and the angle of scanning with folds. The case study produced from CAD has clean, computer-aided images with vector graphics. The possible identified challenges are the parameters (scale and angle) of CAD hatching, angled and curved structures, and the irrelevant shaded regions of other floors and furniture. The BIM dataset is the fuzziest, containing background hatching of additional information like indoor and outdoor shades, thin material layers, and incomplete hatching of some elements. Even though the images are produced using CAD technologies, the ”line clutter” may affect the automated detection.

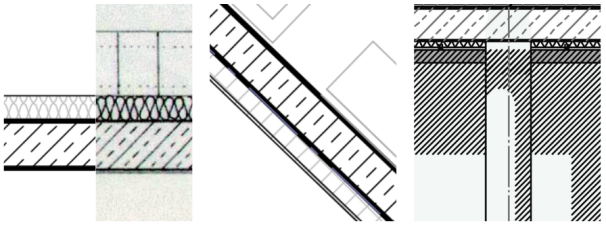

Figure 2: Different hatching representations depending on use cases - a) hand- & CAD-drawn, b) CAD c) BIM.

Initially, the drawings are thoroughly examined to identify the presence of materials, their conventions and any specific patterns used. The hatchings indicating materials vary between computer-aided and hand-drawn drawings and are influenced by the details provided in the legend or key. A challenge arises in interpreting these hatchings, as each data set might use different symbols for specific materials.

Figure 2 a) represents a good example where similar structures from CAD and hand-drawn data sets are compared. Although soft insulation follows the same hatching, the CAD pattern is not centrally aligned, which contrasts with the manual version. In such cases, two classes of material are created - one addressing CAD and one hand-drawn separately.

Further complications with alignment may occur in the context of structures small in thickness and the chosen hatching scale and angle, as depicted in Figure 2 b). When the hatching angle coincides with the structure's angle, interpreting the hatching becomes significantly challenging. The hatching according to the material enables the deep learning model to learn the context of specific materials. In this instance, the likelihood of an insulation layer's presence is considerably high in certain scenarios, as insulation typically runs parallel to load-bearing materials.

Another potential challenge in material detection in architectural drawings lies in the contextual usage of hatching. Common hatching styles might be employed to convey various types of information, such as shading or non-relevant floors, in the context of BIM and CAD-based drawings, which can lead to significant misinterpretations. For instance, in Figure 2 c), the hatching intended to represent shade is identical to that used for a column (specifically masonry). Such scenarios are particularly critical and necessitate meticulous mask preparation to accurately delineate and mark irrelevant information.

Material class definition and annotation

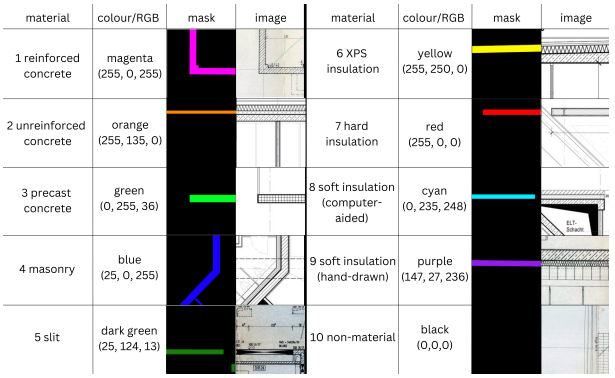

Figure 3: The defined classes for the identified materials

Among the three data sets, 10 classes are identified, as shown in Figure 3. Each class refers to a specific material, the Red Green Blue (RGB) pixel mask with the original image. The class for slits is incorporated as they occupy a significant portion of the wall area. Omitting them from detection could result in considerable volume discrepancies.

Data pre-processing

To prepare the data set, the material classes are defined as masks using GIMP software.These masks are then exported, ensuring they match the original drawing dimensions. Subsequently, both the initial images and their corresponding masks are divided into 224x224 pixel segments. This segmentation includes padding to prevent the loss of any drawing parts, ensuring the entire data set is accurately represented and ready for further processing and analysis.

After producing the patches the data sets are divided into three sets, such as training (80%), validation (10%), and testing (10%). The images are randomly allocated with corresponding masks. Considering the differences between case studies, three CNN models are tested on each case study separately and also all together. The testing on the mixed data set can be used to assess the adaptability of the network to different types of drawings.

To make the CNN model more robust to the changing hatchings regarding the scale and rotation of images, the data set is extended by the same augmented images. The transformations include rotation of 90, 180, and 270 degrees, zooming in and out for 20% and 40%. Each image with the relevant mask is transformed with random rotation and zoom parameters.

Semantic segmentation and fine-tuning

For the case studies, we use three base CNN models, such as MobileNetV2, ResNet50, and U-Net (Sandler et al.,

2019; Ronneberger et al., 2015; He et al., 2015). The architecture setup is conducted using the TensorFlow ML platform with the Keras library. The training is conducted with a learning rate of 0.0001, batch size of 1 and 100 epochs.

To perform more efficient fine-tuning, a new metric, F1 score, for assessment is introduced. The sparse categorical cross-entropy loss function, commonly used for classification tasks, is applied. To improve the model performance for the segmentation task specifically, dice loss is tested at the fine-tuning stage, as it specifically addresses class imbalance by measuring relative overlap. It evaluates an overlap between predicted and true images. The two losses are combined for two goals: firstly, classifying each pixel correctly and secondly, improving spatial overlap between predicted and true masks.

The learning rate is one of the most critical hyper-parameters, as it adjusts the weights of the model after each epoch, directly affecting the training process. To identify the most effective value for this parameter, the Keras Tuner library is used. The options tested for the Adam optimizer include 1e-2, 1e-3, 1e-4, and 1e-5. The algorithm runs the model through five trials, varying the learning rate, and outputs the value that demonstrates the highest Mean IoU value. Considering that the Adam optimizer is used in studied CNN models, square root scaling is applied to test the model with different hyper-parameters. The 4, 16, and 64 batch sizes with relevant learning rates are tested and compared.

To address class imbalance, weights are introduced by using a custom loss function that utilizes weighted sparse cross-entropy. The weights are computed and applied to the cross-entropy loss between the input and target. This approach penalizes false predictions of less represented classes more severely than those of more represented classes.

Overfitting emerges as an issue due to the noise present in the mixed data set, leading the network to learn specific patterns but diminishing its ability to generalize effectively. Combinations of different levels of L2 regularization values with dropout values are tested to help the CNN model generalize better. The leading parameters in assessing the performance of these strategies are stable growth of validation Mean IoU metric and decreasing loss function. The proposed fine-tuning workflow aims to give the CNN model optimal performance, ensuring stable training and effectivity in predictions. All CNN model variations are compared and the decision is made based on the validation metrics.

Results and Discussion

This section first discusses the results of the semantic segmentation differentiating between the data sources. Next, the results of the CNN fine-tuning are described. Finally, we discuss the limitations of this work.

Comparison of CNN models for Semantic Segmentation

The three CNN models are tested on four data sets, including the three data-source-specific ones and the mixed one, and they all show different performance levels in various parameters. Quantitatively, each CNN model's OA and Mean IoU metrics are compared to assess their capability for material identification, depending on the training data set.

In the case of the CAD data set, U-Net, ResNet50 and MobileNetV2 achieve 99.70%, 98.87%, and 99.13% in OA, respectively. BIM case study has a lower identification percentage - 97.48% with U-net, 96.38% with ResNet50, and 97.82% with MobileNetV2 architecture. The hand-drawn data set shows the best results with 99.61%, 99.00%, and 99.11% with U-Net, ResNet50 and MobileNetV2, respectively. Overall, the OA metric achieves more than 96% across all scenarios, largely due to the prevalence of black background pixels.

Table 1: Validation results: Mean IoU

Case Study	**U-Net**	**ResNet50**	**Mobile-NetV2**
Hand-drawn	41.00%	30.16%	39.06%
CAD	77.33%	27.87%	54.34%
BIM	32.58%	23.66%	35.39%
Mixed	30.94%	18.87%	38.87%

Table 1 summarizes the performance of three CNN models trained on four datasets - three of which comprise each case study separately and mixed one combines all three to assess the adaptability of architectures. Comparing the CNN models, the ResNet50 performs the poorest in the case of all data sets. U-Net reaches 77.33% and 41.00% in CAD and Hand-drawn data sets, respectively, being the best performer across the CNN models in these cases. In the BIM case study data set, MobileNetV2 shows the highest performance of 38.87%. Training on mixed data set results in mean IoU values ranging from 18.87% for ResNet50 to 38.87% for MobileNetV2.

Better performance of the models on the CAD and hand-drawn data sets suggests that these drawings' simplicity and reduced contextual complexity facilitate the CNN's ability to learn and generalize features effectively rather than in the BIM data set. This high accuracy in CAD and hand-drawn data sets can be attributed to the fact that there is less noise in the drawings, as they contain less contextual information that might otherwise complicate the learning process. The drawings in these two data sets do not contain the background information and shades as in the BIM data set. The hand-drawn data set performs well despite the noise from scanning.

The mixed data set, which includes more imbalanced information, also demonstrates robust performance with U-Net and MobileNetV2 architectures, highlighting the

adaptability and efficacy of the models in diverse scenarios. ResNet50 architecture shows significantly lower prediction probabilities in comparison.

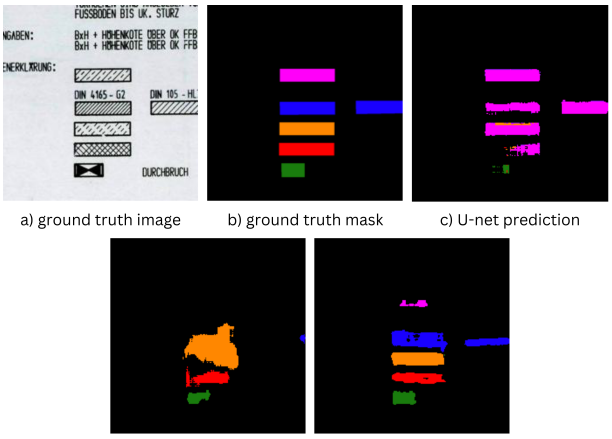

Figure 4: Predictions of the image from the Hand-drawn Case Study trained on the mixed dataset

In visual assessment, in most cases, Figure 4, the MobileNetV2 model consistently predicts the shapes and identifies all present colours, supported by the higher Mean IoU values in the results, Table 1. U-Net, is often precise in shape but correctly predicts mainly pink, reinforced concrete, class, which might be a sign of overfitting. ResNet50 is often inefficient in shape, producing jagged edges or missing shapes. The predictions are tightly connected with metrics. For instance, in the mixed dataset predictions, ResNet50 identifies with the highest probabilities the background and reinforced concrete materials and the images illustrate the successful prediction of mainly these materials.

The confusion matrix presented in Figure 5 visualizes the actual vs. predicted classification classes of the MobileNetV2 CNN network trained on the mixed data set. The previous results are supported, showing the highest true identification of background (99%) and reinforced concrete pixels(68%). The least identified are CAD soft insulation with slit elements.

Material-wise the probabilities of prediction reinforced concrete and background pixels are the highest in most of the cases, which can be accounted for their prevalence in the training images. To the most misidentified materials belong precast concrete, hand-drawn insulation, and slits due to their low presence and complex presentation. Slits represented in the dataset vary largely in size and precast concrete may not be clearly seen in some cases due to the dashed appearance or thin thickness.

The leading factor in choosing the best-performing network is the Mean IoU metric on the mixed data set, as adaptability of the model to the various classes is essential in the material detection task. Table 1 shows that MobileNetV2 reaches the highest percentage in the case of mixed dataset - 38.87%. Hence, it is considered the most adaptable CNN model The trained architecture detects the

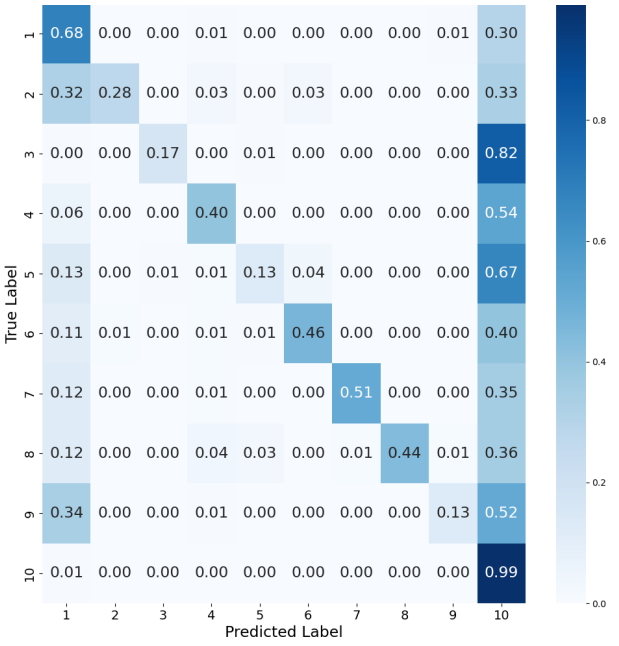

Figure 5: Confusion Matrix of MobileNetV2 trained on the mixed data set: 1-reinforced concrete, 2-unreinforced concrete, 3-precast concrete, 4-masonry, 5-soft hand-drawn insulation, 6-hard insulation, 7-existing, 8-soft CAD insulation, 9-slits, 10-background

colours and shapes consistently with some inaccuracies. MobileNetV2 is chosen for optimization, as it shows the best results in predicting different classes and consistent shapes supported by metrics and visual assessment of predictions.

MobileNetV2 model fine-tuning

The chosen MobileNetV2 CNN model is trained with combined loss, and the F1 Score is added to the evaluation of the model in the trial runs.

The optimization of the learning rate using the Keras tuner is done by assessing the development of 0.01, 0.001, 0.0001, and 0.00001 learning rates over the epochs. 0.01 and 0.001 learning rates plateau on epochs 11 and 12 with only 9.64% Mean IoU, respectively. 0.00001 learning rate reaches the highest number of epochs - 85, but results only in 35.66% in Mean IoU The 0.0001 learning rate yields the best results with 37.59% in Mean IoU.

Using the square root scaling, the CNN model is tested in four trials with different batch sizes and relevantly decreasing learning rates. The trials are done with 50 epochs of training to reduce the time for running the trials. It is enough to see the main training tendencies of architectures. The batch size of 1 is leading in both metrics over other setups. The metrics seem to decrease with an increase in batch size and learning rate. The reason for seen behaviour might be the present class imbalance and high diversity of images. The more granular updates help to reduce overfitting to the patterns present in the data set. This indicates that the CNN model processes more individual examples per epoch, allowing rare classes to contribute more significantly to the gradient—a crucial factor

given the dominance of two prevalent classes in the current setup.

When weighted balancing of classes is introduced with weighted sparse cross-entropy loss function, the weighted model performs 2% and 0.4% lower in Mean IoU and F1 Score, respectively. The weighted CNN model seems to not learn the reinforced concrete class, as its weight penalty is the highest. At the same time, it predicts the other material colours inconsistently, such as hard insulation and unreinforced concrete. This can be the effect of instability due to highly different weights. Some classes appear very rarely and the CNN model aims to improve on those when they are not even present, hence the weights unbalance the training.

With the addition of augmented data to the training data set, the drop in metrics is shown. The Mean IoU drops by 12.31%. The underlying reasons for the decrease in performance might be first, the quality depletion due to zooming effects, where hatchings become less identifiable by the network, and second, the distortion of spatial relationships, as the integrity of spatial orientation of elements is high in type of drawings like sections. Mainly the added noise might bring unfavourable behaviour of the neural network.

Visualisation of the loss over the epochs showcases signs of overfitting. The training loss decreases over the epochs but the validation loss increases, which is abnormal behaviour and is considered an overfitting of the CNN model. On the opposite, the metric of Mean IoU, despite the increasing validation loss, increases as well. Overfitting behaviour is tackled by the common strategies of addressing it - L2 regularization and dropout. However, high amplitude oscillations are observed which can be the cause of high noise, introduced by dropout rates and dampening of the weights. As the favourable behaviour of validation loss drop is shown, the metric of Mean IoU drops, which can be a result of slower learning and exacerbation of existing problems of class imbalance and noise. The high instability and amplification of complex data sets result in persistent overfitting or low and oscillating performance when overfitting is reduced.

The combined loss function brought down the important metric of Mean IoU over 100 epochs at first. This can be accounted for by the dice loss focusing more on the overlapping regions, rather than pixel-wise classification. Considering the context, the variability of the data set, and class imbalance, the learning may struggle to optimize the Mean IoU. However, with the additional 50 epochs, the CNN model with combined loss started to perform slightly better, specifically with improvement in the detection of under-represented materials like precast concrete, masonry, CAD insulations and slits. We can visually identify an improvement in the prediction in Figure 6. We conclude that the CNN model with one loss, sparse categorical cross-entropy, function learns more the most represented classes, being reinforced concrete and background, while in combination with dice loss, gradual improvements over other classes are observed. The second case improved over

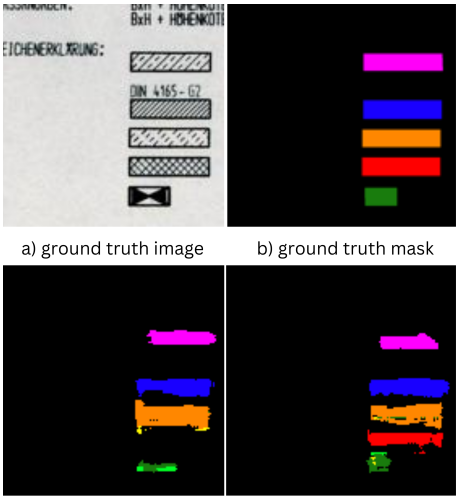

Figure 6: Comparison of predicted image by original MobileNetV2 and the fine-tuned CNN model

more epochs, due to the stabilization of gradients and better at handling class imbalance, particularly in problematic classes. Dice loss inherently gives more weight to underrepresented classes (since it focuses on overlap), which can lead to higher IoU in these classes after extended training.

Limitations

A notable challenge in the training process is the overfitting of dominant classes, such as reinforced concrete and non-material pixels and underfitting of those that are underrepresented, such as soft insulation, and slits. This issue arises from the high-class imbalance within the data set, leading to an overfitting tendency accompanied by increasing validation loss over successive epochs. The model demonstrates a low probability of accurately identifying less common elements such as slits, hand-drawn insulation, and unreinforced concrete, indicating a propensity to either misclassify rare hatch patterns as more prevalent ones or overlook them entirely.

One of the most critical limitations of the proposed methodology is the high dependency of the CNN model on the quality, consistency, and variation in the training data sources. The model's robustness is notably diminished when identifying materials that are poorly represented in the data set, making it less effective in diverse or noisy conditions.

These limitations are directly linked to the accuracy of the resulting quantity estimations. The precision of numerical predictions is heavily dependent on the quality of material representation in the data set.

Conclusion and Outlook

This paper introduces a novel approach for automated material detection and classification in architectural drawings using CV. It focuses on the data preparation and semantic segmentation phases with fine-tuning and is based on the Master's thesis work of Kairlapova (2024).

The proposed methodology emphasizes the automated detection of materials in architectural drawings rendered in various stylistic conventions. Among the three tested CNN models, the MobileNetV2 architecture demonstrates the best adaptability to a mixed dataset, achieving a Mean IoU of 38.87%. The implementation across three case studies with three different data sources, such as BIM, CAD, and hand-drawn datasets, illustrates the overall capability of CNN models to learn and generalize hatch patterns. However, this performance is highly contingent on the representation of materials and the complexity of the drawings. The CNN model exhibits a tendency to overfit, resulting in a bias towards the most prevalent classes—namely, the background and the dominant material, reinforced concrete.

Of the fine-tuning techniques applied, only the combined loss function, integrating sparse categorical cross-entropy with dice loss, yields a marginal performance improvement, with a Mean IoU of 38.91% after additional epochs. Overfitting remains a critical issue, warranting further investigation. Overall, the methodology demonstrates potential in leveraging ML for material detection, with possible applications in the reconstruction of as-built models and the development of MPs, as it effectively identifies the most represented, structural materials.

Future work should focus on improving the model's robustness and accuracy through larger training data, enhanced training techniques, better representation of complex geometries, and more sophisticated methods for handling ambiguous or visually similar materials. One of the examples is the usage of synthetic data generation to create more diverse training data sets, helping the model distinguish between similar materials. Generally, the additional data for training is the best way to see the improvement in the results. In addition, instead of utilising one CNN model for the whole task, tasks of detection and semantic segmentation can be done with separate CV technologies.

References

Carrara, A., Tee, L., Nousias, S., and Borrmann, A. (2024). Deep learning-based segmentation and 3d reconstruction for heterogeneous mixed-use building layouts. In Proceedings of the 31st International Workshop on Intelligent Computing in Engineering.

Darko, A., Chan, A. P., Adabre, M. A., Edwards, D. J., Reza, H. M., and Ameyaw, E. E. (2020). Artificial intelligence in the aec industry: Scientometric analysis and visualization of research activities. Automation in Construction, 112:103081.

European Commission (2020). A Renovation Wave for Europe - greening our buildings, creating jobs, improving lives.

García, J. C., Caro, D., Foster, G., Pristerà, G., Gallo, F., and Tonini, D. (2023). Techno-economic and environmental assessment of construction and demolition waste management in the European Union - Status quo and prospective potential.

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition.

Honic, M., Kovacic, I., Gilmutdinov, I., and Wimmer, M. (2020). Scan to BIM for the Semi-Automated Generation of a Material Passport for an Existing Building. pages 338–346. Eduardo Toledo Santos and Sergio Scheer.

ISO 128-3:2022 (2022). Technical product documentation (TPD) — General principles of representation — Part 3: Views, sections and cuts. ISO (International Organization for Standardization).

Kairlapova, A. (2024). Material detection and classification in 2D architectural drawings using Computer Vision for Material Passports. Master's thesis, Technische Universität München.

Kingma, D. P. and Ba, J. (2017). Adam: A method for stochastic optimization.

Mafipour, M., Ahmed, D., Vilgertshofer, S., and Borrmann, A. (2023). Digitalization of 2d bridge drawings using deep learning models. Publisher Copyright: © 2023 30th EG-ICE: International Conference on Intelligent Computing in Engineering 2023. All rights reserved.; 30th International Conference on Intelligent Computing in Engineering 2023, EG-ICE 2023 ; Conference date: 04-07-2023 Through 07-07-2023.

Neufert, E., Kister, J., Lohmann, M., Merkel, P., and Brockhaus, M. (2019). Bauentwurfslehre : Grundlagen, Normen, Vorschriften über Anlage, Bau, Gestaltung, Raumbedarf, Raumbeziehungen, Maße für Gebäude, Räume, Einrichtungen, Geräte mit dem Menschen als Maß und Ziel ; Handbuch für den Bauchfachmann, Bauherrn, Lehrenden und Lernenden. Springer Vieweg, Wiesbaden.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation.

Russell, S. and Norvig, P. (2021). Artificial intelligence, global edition a modern approach. Pearson Deutschland.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. (2019). Mobilenetv2: Inverted residuals and linear bottlenecks.

Xiao, Y., Chen, S., Ikeda, Y., and Hotta, K. (2020). Automatic recognition and segmentation of architectural elements from 2D drawings by convolutional neural network. Proceedings of the 25th International Conference on Computer-Aided Architectural Design Research in Asia, CAADRIA 2020, pages 843–852.