| Model | Strategy | Target variable | Combination method |
|---|---|---|---|
| Model 1 | A: Multi-kernel learning | Price | Linear weighting |
| Model 2 | B: Multi-view concatenation by feature extraction | Price | Concatenation |
| Model 3 | B*: Multi-view concatenation by boosting | Error term | Boosting |
| Model 4 | C: Hybrid multi-view neural network | Price | Fusion by concatenation |
| Model 5 | C*: Multi-view neural network | Price | Fusion by concatenation |

Table 2: Summary of the different models compared.

used was not entirely complementary (in other words: overlapping) to the housing attributes, which also contain location information. In the work of **?** (**?**), visual aesthetics extracted from the interior images might have improved predictive performance, because these features were not captured by the relational attributes. Second, using not only an independently and identically distributed hold-out set, but an out-of-sample test set further increases the required maturity level of the algorithm. On the in-sample validation dataset, the performance improved on RMSE and MAE. Nevertheless, on the out-of-sample predictions, only a lower RMSE could be reported (Table 3). As the RMSE penalizes larger errors stronger, it seems that extreme deviations are reduced. Though, the higher MAE could be caused by noise introduced by CNN's prediction of the residual. Finally, models 4 and 5, both based on the multi-view neural network strategy, performed best by far. Model 4, the interpretable hybrid model provided 7.6% more accurate predictions in terms of MAE. Model 5, the more complex black-box model, outperformed the baseline by 13.4% in MAE. In the next section we show the possibilities to interpret models 3 and 4.

### 4.2 Interpretability of Multi-View Real Estate Appraisal Models

We report a selective set of coefficients of model 3 and model 4 in Table 4. While the constants differ between the models, the estimated effect of the size measured in square feet, the amenity fireplace or the location in Asheville have approximately the same size in both models. For example, the house price increases by approximately 5% when the house has a fireplace, ceteris paribus. As the continuous variables are min-max normalized, inferential statements cannot be made. Nevertheless, the sign and size of the coefficients show the direction and strength of impact on the house price. For example, the condition unsound is associated with a negative price influence, while having additional half-bathrooms adds value to the real estate. From interpreting the coefficients magnitude, we can conclude that location aspects captured by the satellite image have a strong influence on the appraisal value, as the satellite image is among the top-3 largest coefficients for model 3 and top-15 for model 4. The difference between the magnitude of the image data coefficient of model 3 and 4 can be explained by the different scales of the variables. While model 3 includes the impact of the satellite image by the predicted residual and is thus measured in log-USD, model 4 learned the impact of the image implicitly. When the predicted residual changes by 1%, the house price changes by 1.05% (model 3). Furthermore, model 3 has the advantage that besides the coefficients, also standard errors, t-values, and significance levels can be extracted from the linear regression model. Concluding, the derived coefficients and statistical measures can enhance the model understanding and transparency.

In summary, we can conclude the following lessons learned from our experiments (Table 5): Models using multi-kernel learning (Strategy A) are easy to train, however their performance seems to depend strongly on the weighting of the kernels. We suggest using these models as an advanced baseline, in addition to classic hedonic pricing models. Furthermore, models based on multi-view concatenation (Strategy B) seem to reliably increase predictive performance. This strategy is easier to optimize compared to multi-view neural networks, however, selecting the right (intermediate) target seems to be essential for successful learning. Moreover, it seems that the boosting approach of **?** (**?**) (Strategy B*) is more beneficial when the additional view is largely complementary to the other view. Nevertheless, the strength of this approach lays in interpretable coefficients and a statistical measurement of the effects. Consequently, this strategy seems very suitable for research contexts. From our experience, multi-view neural networks (Strategy C and C*) perform best. It seems that learning a latent subspace leads to an effective feature representation even if the multiple

| Model | MAE | RMSE |
|---|---|---|
| Baseline | 40,303 | 71,518 |
| Model 1 | 49,019 | 78,983 |
| Model 2 | 38,395 | 61,663 |
| Model 3 | 43,173 | 68,362 |
| Model 4 | 37,225 | 61,429 |
| Model 5 | 34,890 | 56,099 |

Table 3: Performance of the different models. Before evaluation, the log-transformation is inversed to gain better interpretability.

| Variable | Model 3 | Model 4 |
|---|---|---|
| Constant | 11.07 | 0.27 |
| Square Feet | 0.98 | 1.00 |
| Year Built | 0.34 | 0.53 |
| Half Bathrooms | 0.12 | 0.12 |
| Fireplace | 0.05 | 0.05 |
| Style 1.5 Conventional | 0.05 | 0.04 |
| Locality Asheville | 0.17 | 0.11 |
| Condition Unsound | -0.23 | -0.33 |
| Satellite Image | 1.05 | 0.41 |

Table 4: Selective set of coefficients of model 3 and 4

| Strategy | Benefit | Disadvantage | Suggested use |
|---|---|---|---|
| A | Low training complexity | Lower accuracy | Advanced baseline model |
| B | Sequential training process | Difficult target variable selection | Alternative for strategy C |
| B* | Statistical interpretation | Complementary views required | Research purposes |
| C | Interpretability | High complexity of training | Multi-use |
| C* | Performance | High complexity of training | Predictive model |

Table 5: Summarization of benefit and disadvantage of the different strategies

views overlap in some features. In our experiment, the location of the house is, in addition to the satellite image, also partly captured by city dummy variables. On the downside, according to **?** (**?**), it can become difficult to achieve convergence of the neural network model, which raises the complexity of training. In addition, model 5 (Strategy C*) comes as a black-box model. To mitigate this weakness, a semi-transparent multi-neural network as suggested by **?** (**?**) can enhance interpretability. Nonetheless, it comes with an interpretability-accuracy trade-off, as model 4 has weaker predictive performance than model 5. We suggest using the multi-view neural networks for predictive models in applications where explainability is not a key objective.

## 5 Conclusion

Of course, our work is not without limitations. We have not yet investigated the interrelations between the structured housing attributes and the satellite images. Future research could, for example, analyze which information is overlapping between views and which information can only be captured by one or the other view. Another limitation is related to our data source. We use only a single dataset of Asheville, NC to assess the effect of the learning strategy on predictive performance. As related literature typically refers to neural network architecture search, future research should perform replication and ablation studies to examine, if results can be reproduced across datasets, image types and domains (beyond housing). Despite these limitations, the following implications for research and practice can be derived from our findings. Satellite images can clearly improve the accuracy of computer-assisted mass appraisal. Our results indicate that the MAE can be reduced by up to 13%, depending on the chosen multi-view learning strategy. Therefore, banks and lenders should consider using visual data to improve their real estate appraisal estimates. Moreover, different techniques match different purposes and user groups. Researchers interested in a statistical interpretation of the results might find the boosting strategy of model 3 more appealing. Practitioners mainly interested in predictive accuracy might prefer model 5, the multi-view neural network.

Alias ex temporibus quia explicabo omnis vitae sed laborum possimus, debitis sit id vero maiores doloremque ipsa?Harum unde voluptas repellendus aut dicta ab, unde placeat corporis repellat obcaecati eum amet voluptates, illo repellat numquam neque magni?Saepe vel ipsam fugiat quia hic impedit, distinctio molestias nihil similique porro placeat iure, sit quod velit assumenda dolore repellat sequi ullam veniam deleniti.Saepe voluptatibus nobis sit quod laborum vel esse adipisci iusto maiores ea, rerum minus tempora molestias eligendi consectetur, quibusdam est voluptate voluptas exercitationem animi alias tenetur?Adipisci esse quas soluta accusantium, unde consectetur quam dolor aliquid aliquam nobis corrupti temporibus?Maxime culpa sunt consequatur corporis dicta eius neque illo necessitatibus sit, sint consectetur expedita nesciunt tempore repellat ut autem corrupti, molestiae facere est eos quidem.Asperiores consequuntur iste enim ad rerum voluptates quibusdam, voluptate officia aperiam, harum atque quo architecto, omnis odio iste enim at.Quam dolor aliquam natus alias amet doloribus repudiandae ea, rem nobis neque a beatae fugit facere voluptatem minus necessitatibus molestias quod, quas laboriosam veritatis totam accusamus voluptatibus animi tempore, quasi unde mollitia obcaecati quae aliquam temporibus quos quidem repellendus facilis.Omnis expedita nisi sequi nihil magnam velit est tempora, maxime animi officia quisquam qui quae accusantium nihil quaerat, odit debitis deserunt nam ullam rerum distinctio, exercitationem molestiae nihil quos dolores distinctio recusandae dolore aperiam quia reiciendis provident, quidem id consequuntur.Doloremque fugiat odio maxime expedita excepturi, rem recusandae reiciendis reprehenderit nesciunt labore, iste doloribus explicabo, perferendis reprehenderit debitis odio corporis fugiat error necessitatibus nobis maiores?Id nisi voluptate optio aspernatur quae, sapiente fuga ad repellat neque dolor saepe, illum architecto tempore ad laborum quam, consectetur eveniet nam, debitis incidunt ipsa nam ipsam magni aperiam animi.Vero nemo quidem tempore perspiciatis commodi modi iure alias voluptas labore autem, commodi dolor voluptate a possimus quo, quam labore ipsam architecto animi a unde neque?Ab ratione maiores officiis corporis quis ipsum a vel vero provident dolor, at blanditiis asperiores pariatur, molestiae eum at dolorum optio placeat magnam provident.Expedita harum quo eligendi illo quia rem dolor, ex praesentium veniam non nisi voluptates molestias ullam autem optio blanditiis?Quo eos voluptatibus sequi, non fugit obcaecati nemo optio incidunt recusandae illum asperiores accusamus, voluptatibus quas itaque consequatur provident fuga quisquam, eligendi magni molestias perspiciatis, reiciendis aut eligendi obcaecati expedita?Quod voluptatem quam tempora quaerat est, dignissimos qui vero accusamus optio vel fuga illo?Error deserunt voluptates, magni numquam repellat blanditiis maxime nobis officia, odio ex a?Odio suscipit eveniet beatae inventore illum exercitationem molestias harum quisquam unde, ipsam illo aut ullam harum provident, unde modi molestias sequi, repellat quia ex pariatur eaque magni qui error, placeat eligendi eveniet nisi animi neque accusamus unde commodi.Quasi soluta eligendi et quod porro,