




What a Comfortable World: Ergonomic Principles Guided Apartment Layout Generation

Piotr Nieciecki^{†1} , Aleksander Plocharski^{†1,2} , and Przemyslaw Musialski³ 

¹Warsaw University of Technology, Poland

²Akces NCBR, Poland

³New Jersey Institute of Technology, United States

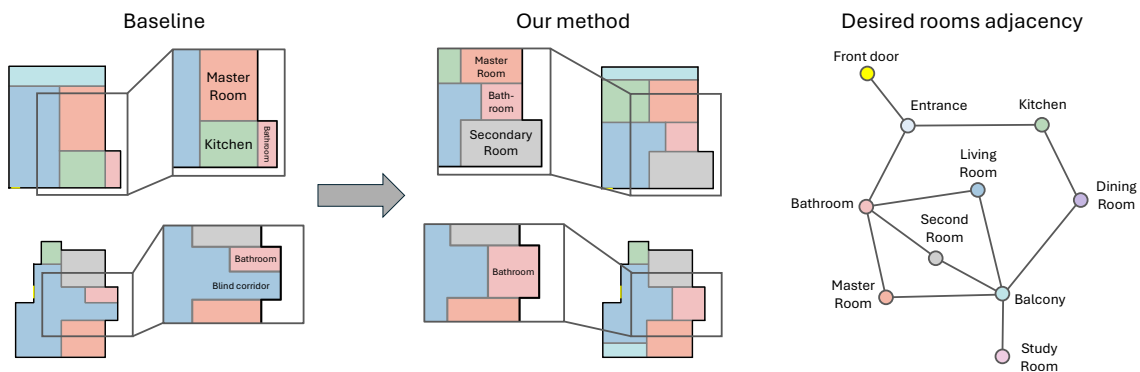


Figure 1: The left side illustrates representative examples of ergonomic floor plans generated by the proposed method, with comparison to a baseline method. The right panel displays the adjacency graph defining the desired spatial proximity between specific room pairs.

Abstract

Current data-driven floor plan generation methods often reproduce the ergonomic inefficiencies found in real-world training datasets. To address this, we propose a novel approach that integrates architectural design principles directly into a transformer-based generative process. We formulate differentiable loss functions based on established architectural standards from literature to optimize room adjacency and proximity. By guiding the model with these ergonomic priors during training, our method produces layouts with significantly improved livability metrics. Comparative evaluations show that our approach outperforms baselines in ergonomic compliance while maintaining high structural validity.

CCS Concepts

• **Computing methodologies** → *Shape modeling; Neural networks*; • **Applied computing** → *Architecture (buildings)*;

1 Introduction

Automated generation of residential floor plans is a long-standing challenge in computer graphics, aiming to support architects with rapid exploration of design variations. Recent data-driven approaches, particularly those leveraging transformer architectures [VSP* 17], have shown remarkable ability to learn complex spatial distributions from large datasets like RPLAN [WFT* 19].

However, a fundamental limitation of pure data-driven methods

is their reliance on the quality of training data. As observed in recent work on furniture arrangement [LGWM22], real-world datasets often contain layouts that are geometrically valid but potentially suboptimal regarding circulation or room placement standards. A model trained naively on such data faithfully reproduces these inefficiencies. While Leimer et al. [LGWM22] addressed this by injecting differentiable guidance into the generation of discrete object arrangements, extending this paradigm to *structural floor plan generation* presents a distinct challenge. Unlike moving furniture within a fixed container, generating a floor plan requires optimizing the container itself—defining the topology, adjacency, and tiling of rooms .

[†] Equal contribution

In this work, we propose a method to generate residential layouts with improved architectural compliance by integrating domain-specific principles directly into the generative process. We formulate a differentiable loss function based on architectural standards [NN12]. By incorporating these architectural adjacency priors into a GPT-2 based architecture [RWC*19] via a dynamic weighting scheme, we guide the model to produce layouts that adhere more closely to design guidelines than the training distribution.

Our contributions are: (1) a guideline-derived proximity cost formulation adapted for Manhattan-world room polygons; (2) generalization of differentiable loss integration presented in [LGWM22] and (3) quantitative evidence that our method improves compliance metrics on the RPLAN dataset while maintaining high parsability, albeit with modest trade-offs in area coverage.

2 Related Work

Early layout synthesis relied on constraint satisfaction or evolutionary algorithms [MCP02], but the field has shifted toward deep generative models. Graph-based approaches like constraint graphs [PGK*21], House-GAN++ [NHC*21] and HouseDifusion [SHF22] represent layouts as nodes and edges, ensuring topological validity but often struggling with precise geometry. Most recently, transformer-based autoregressive models have become the state-of-the-art for vector graphics generation. SceneFormer [WYN20] and FaçAID [PSPSM24] demonstrate that transformers can effectively sequence geometric tokens to generate indoor scenes and building facades, respectively. We build on this sequence-based representation, adapting the GPT-2 architecture [WDS*20] to predict room polygons as coordinate tokens.

Purely generative models often fail to capture high-level functional requirements. To address this, neuro-symbolic methods hybridize neural generation with explicit rule-based constraints. LayoutEnhancer [LGWM22] pioneered this for indoor scenes by using differentiable scalar functions to penalize suboptimal furniture arrangements during training. This allowed the model to improve upon its own training data. Our work extends this principle from the *interior content* domain (furniture) to the *structural domain* (floor plans). We adapt the loss formulation to operate on polygon vertices and adjacencies, ensuring that the generated structures align with established architectural circulation guidelines [NN12].

3 Ergonomic cost

Using domain-specific literature [NN12] we determined that minimizing the following distances have a positive impact on the overall comfort of apartment use: (i) the distance from the entrance room to the front door; (ii) the distances from the entrance room, living room, master room and second room to the bathroom; (iii) the distances from the entrance room and dining room to the kitchen; (iv) balcony adjacency to kitchen, dining room, living room, master room, secondary room, or study room.

Taking the above aspects into account, we define a set of room-type-specific cost functions to quantify layout ergonomics.

Entrance cost. For each entrance room polygon $r \in R_{\text{entrance}}$, the

entrance cost is defined as

$$E_{\text{entrance}}(r) = \text{dist}(r, d),$$

where d denotes the front door (represented as a line segment) and dist is the Euclidean distance in meters.

Kitchen cost. For kitchens, we first pair each entrance and dining room with the nearest kitchen present in the floor plan. The cost associated with a given kitchen is then computed as the average distance to all entrances and dining areas assigned to it:

$$E_{\text{kitchen}}(r) = \frac{1}{|A(r)|} \sum_{r' \in A(r)} \text{dist}(r, r'),$$

where $r \in R_{\text{kitchen}}$ is a kitchen and $A(r)$ is a set of rooms assigned to kitchen r .

Bathroom cost. The cost of each bathroom $r \in R_{\text{bathroom}}$ is defined in an analogous manner, but for each entrance room, living room, master room and second room we assign the nearest bathroom:

$$E_{\text{bathroom}}(r) = \frac{1}{|A(r)|} \sum_{r' \in A(r)} \text{dist}(r, r'),$$

Balcony cost. The cost is defined as the minimum distance to any room from the set of preferred adjacent room types (R'):

$$E_{\text{balcony}}(r) = \min_{r' \in R'} \text{dist}(r, r').$$

Overall ergonomic cost is computed as the mean cost over all rooms of applicable types:

$$E = \frac{1}{|R^*|} \sum_{r \in R^*} E_{a(r)}(r), \quad (1)$$

with $R^* = R_{\text{entrance}} \cup R_{\text{kitchen}} \cup R_{\text{bathroom}} \cup R_{\text{balcony}}$ and $a(r)$ is the type of room r .

4 Floor plans generation

Following trends in related work, we employ the GPT-2 model [RWC*19] (specifically the implementation included in the Hugging Face library [WDS*20]).

4.1 Floor plan representation

We encode each floor plan as a token sequence $S = (b, d, r_1, \dots, r_n)$, where b is the boundary, d the front door placement, and r_i the i -th room. Boundary and door appear first; room order may vary.

Each segment starts with a type token (s_b for boundary, s_d for door, s_r^t for a room of type t) followed by quantized vertex coordinates (x, y) at fixed resolution:

$$b = (s_b, x_{b1}, y_{b1}, x_{b2}, y_{b2}, \dots, x_{bk}, y_{bk})$$

$$d = (s_d, x_{d1}, y_{d1}, x_{d2}, y_{d2})$$

$$r_i = (s_i^t, x_{i1}, y_{i1}, x_{i2}, y_{i2}, \dots, x_{il}, y_{il})$$

Following [WYN20, PGK*21, LGWM22, PSPSM24], we augment each token with an xy-index and a vertex-index. The xy-index is 1 for x-coordinates, 2 for y-coordinates, and 0 otherwise. The vertex-index is 0 for start tokens and counts coordinate pairs (1 for the

first vertex, 2 for the second, etc.). Each index has its own learned embedding, added to the token and positional embeddings.

4.2 Ergonomic loss

To inject expert knowledge into the model, we introduce a custom loss function using the ergonomic cost terms described in Section 3.

First, we define a differentiable distance metric between rooms. For room polygons r and s let $V_r = \{p_1^r, p_2^r, \dots, p_{n_r}^r\}$ and $V_s = \{p_1^s, p_2^s, \dots, p_{n_s}^s\}$ be the sets of their vertices. Then the distance metric D is defined as:

$$D(r, s) = \langle e, \text{softmax}(\beta \cdot e) \rangle, \text{ where } e_{ij} = \|p_i^r - p_j^s\|_2.$$

Here, β is a temperature parameter that determines the hardness of the softmax function (our experiments use $\beta = 10$).

Similarly to ergonomic cost, ergonomic loss is defined as a combinations of room-type-specific losses. For each entrance room it is equal to the differentiable distance between the entrance room polygon and the front door polygon.

$$L_{\text{entrance}}(r) = D(r, d),$$

where d is the front door polygon. The mean value over all entrance rooms in a floor plan forms the full entrance room loss term:

$$L_{\text{entrances}} = \frac{1}{|R_{\text{entrance}}|} \sum_{r \in R_{\text{entrance}}} L_{\text{entrance}}(r).$$

For kitchens, each entrance room and dining room distance is calculated to each kitchen. Then a differentiable minimum function is used to determine the distance from the nearest kitchen. The kitchen loss term is a mean value of these distances. Let $r_i \in R_{\text{entrance}} \cup R_{\text{dining}}$, $r_j \in R_{\text{kitchen}}$ and $\Delta_{ij} = D(r_i, r_j)$, then:

$$L_{\text{kitchen}}(r_i) = \langle \Delta_{ij}, \text{softmax}(\beta \cdot \Delta_{ij}) \rangle,$$

$$L_{\text{kitchens}} = \frac{1}{|R_{\text{entrance}} \cup R_{\text{dining}}|} \sum_{r_i \in R_{\text{entrance}} \cup R_{\text{dining}}} L_{\text{kitchen}}(r_i).$$

Loss for bathrooms is defined in an analogical way, but using the proximity of entrances, living rooms, master rooms and second rooms. Let $R'' = R_{\text{entrance}} \cup R_{\text{living}} \cup R_{\text{master}} \cup R_{\text{second}}$, then:

$$L_{\text{bathrooms}} = \frac{1}{|R''|} \sum_{r_i \in R''} L_{\text{bathroom}}(r_i).$$

Balcony loss is similar, but here we are taking a differentiable minimum of proximate distance losses:

$$L_{\text{balconies}} = \langle L_{\text{balcony}}(r_i), \text{softmax}(\beta \cdot L_{\text{balcony}}(r_i)) \rangle, \quad r_i \in R'.$$

Final ergonomic loss is a mean of all possible to calculate losses:

$$L_E = \frac{\sum_a \delta_a L_a}{\sum_a \delta_a},$$

with $a = \{\text{entrances}, \text{kitchens}, \text{bathrooms}, \text{balconies}\}$ and $\delta_a = 1$ if the loss is applicable to the floor plan and $\delta_a = 0$ otherwise.

4.3 Ergonomic loss usage

We integrate our ergonomic loss into the training process using a methodology similar to one presented in [LGWM22]. The ergonomic loss is calculated using ground-truth sequences where

a single token is replaced by a value \bar{v} derived from the model's predicted probability distribution.

Token \bar{v} is equal to the expected value in a small window around the most likely value of the token

$$\bar{v} = \frac{\sum_j \mathcal{N}(v_j | \hat{v}, \sigma) P(v_j) v_j}{\sum_j \mathcal{N}(v_j | \hat{v}, \sigma) P(v_j)},$$

with $\mathcal{N}(x | \hat{v}, \sigma)$ being the normal distribution centered at \hat{v} with standard deviation σ . \hat{v} is equal to the most probable token and $P(v)$ is a estimated by the model. During our experiments σ was equal to $1/\rho$, where ρ is the resolution of floor plans quantization.

Since the ergonomic loss is differentiable w.r.t. the room polygon vertices, the loss value for a given predicted token \bar{v} is computed only when both \hat{v} and its corresponding ground-truth element represent x- or y-coordinate of a room polygon vertex.

During the learning process we are combining the standard cross-entropy loss L_C with our ergonomic loss.

$$L = (1 - \alpha) \cdot L_C + \alpha \cdot L_E,$$

where α is the value of our ergonomic loss of the ground-truth floor plan divided by a scaling factor γ and clamped to $[0; 1]$ range. During our experiments, γ was set to 30 based on ergonomic loss distribution from test samples. This definition increases the influence of ergonomic loss on non-ergonomic data, encouraging the model to rely more on expert knowledge for such cases.

5 Results

We train on RPLAN [WFT*19] (quantized to 256). After validity filtering, the dataset contains 80,405 plans. We use a 90/5/5% train/val/test split. The training set was expanded to ~159,000 examples via augmentation (rotation, symmetry, room permutation). The model (GPT-2, 25 layers, 16 heads, 256 embedding dim, 320 token context, 20M parameters) is trained for 150 epochs. A baseline model with identical configuration but trained with standard cross-entropy loss serves as a baseline. Inference uses greedy decoding. During testing the model starts from ground truth boundaries and doors (as constraints), though from-scratch generation is also possible. Code and additional results are available on <https://comfortableworld.github.io>.

Qualitative results. Figures 1 and 2 illustrate our results. Ergonomic loss minimization is evident in bathroom placement (closer to master/secondary rooms) and reduced blocking of functional rooms (e.g., kitchen blocking bathroom in baseline, Fig. 1).

Quantitative results. We define test metrics: *Parsability* (successfully parsed sequences); *Validity* (no self-intersections); *Fully covered* (complete interior coverage); *No room overlapping*; *Ergonomic cost* (see Eq. 1, lower is better); and *Perfect ergonomic cost* (% with 0 cost). All metrics, besides ergonomic cost are binary. Table 1 compares these metrics. Our model outperforms the baseline on ergonomics while maintaining comparable generation quality.

6 Limitations and Conclusions

Limitations. While our method successfully optimizes ergonomics, we observe a trade-off in geometric packing efficiency,

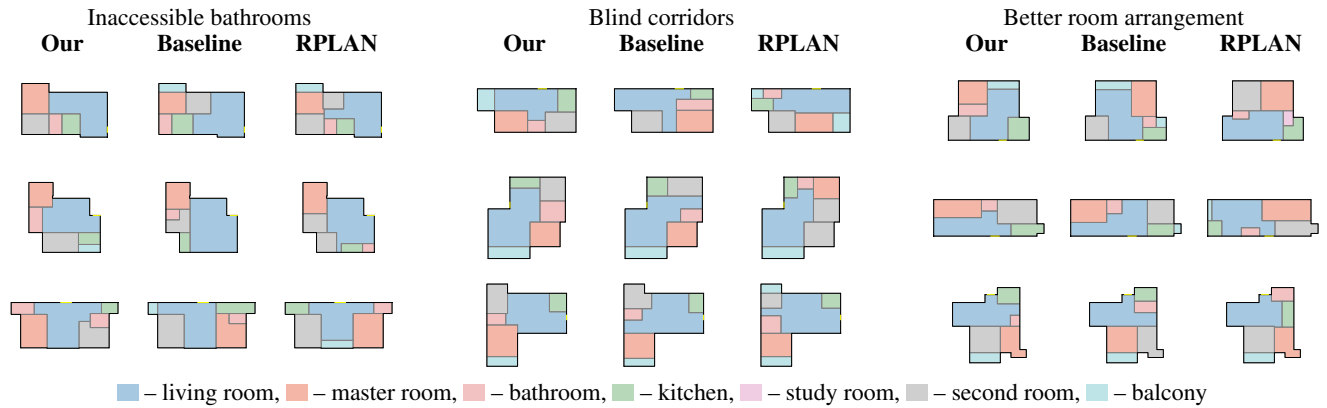


Figure 2: Qualitative comparison of results between baseline, our method and samples from RPLAN dataset. Reduction of inaccessible bathrooms and blind corridors compared to the baseline; the final column shows improved room arrangements.

resulting in slightly lower area coverage compared to the baseline. Additionally, the current model operates unconditionally, lacking explicit constraints for room counts or types. Finally, rare logical inconsistencies persist, such as entrance doors opening into spaces like balconies.

Conclusions. We presented a transformer-based method for residential floor plan generation that integrates domain-specific ergonomic principles. By incorporating differentiable loss functions for room adjacency, our model outperforms the baseline in ergonomic metrics despite the limitations of the training data. This approach demonstrates that incorporating domain-specific expert knowledge into training can effectively guide generative models toward more architecturally compliant and practical designs.

As future work, we plan to add conditional control over room types and counts, improve geometric packing and logical consistency, and experiment with bigger, more recent model architectures.

7 Acknowledgments

This research was carried out with the support of the High Performance Computing Center at Faculty of Mathematics and Information Science Warsaw University of Technology.

Generative AI tools were used for brainstorming and polishing the manuscript text. All scientific content and analysis were produced entirely by the authors.

Table 1: Quality metric values for test results of both the baseline and our full method as well as ground truth RPLAN dataset metrics.

Metrics	Baseline	Our	RPLAN
Parsability (%)	99.93	99.93	100
Validity (%)	96.98	97.71	100
Fully covered (%)	79.83	74.19	100
No overlapping (%)	98.02	97.76	100
Ergonomic cost (m)	0.502	0.353	0.509
Perfect ergonomic cost (%)	34.32	43.75	28.07

References

- [LGWM22] LEIMER K., GUERRERO P., WEISS T., MUSIALSKI P.: Lay-outenhancer: Generating good indoor layouts from imperfect data. In *SIGGRAPH Asia 2022 Conference Papers* (2022), pp. 1–8. 1, 2, 3
- [MCP02] MICHALEK J., CHOUDHARY R., PAPALAMBROS P.: Architectural layout design optimization. *Engineering Optimization* 34, 5 (2002), 461–484. doi:10.1080/03052150214016. 2
- [NHC*21] NAUATA N., HOSSEINI S., CHANG K.-H., CHU H., CHENG C.-Y., FURUKAWA Y.: House-gan++: Generative adversarial layout refinement network towards intelligent computational agent for professional architects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021), pp. 13632–13641. 2
- [NN12] NEUFERT E., NEUFERT P.: *Architects' data*. John Wiley & Sons, 2012. 2
- [PGK*21] PARA W., GUERRERO P., KELLY T., GUIBAS L. J., WONKA P.: Generative layout modeling using constraint graphs. In *Proceedings of the IEEE/CVF international conference on computer vision* (2021), pp. 6690–6700. 2
- [PSPSM24] PLOCHARSKI A., SWIDZINSKI J., PORTER-SOBIERAJ J., MUSIALSKI P.: FaçAID: A Transformer Model for Neuro-Symbolic Facade Reconstruction. In *SIGGRAPH Asia 2024 Conference Papers* (2024), SA '24. doi:10.1145/3680528.3687657. 2
- [RWC*19] RADFORD A., WU J., CHILD R., LUAN D., AMODEI D., SUTSKEVER I., ET AL.: Language models are unsupervised multitask learners. *OpenAI blog* 1, 8 (2019), 9. 2
- [SHF22] SHABANI M. A., HOSSEINI S., FURUKAWA Y.: Housediffusion: Vector floorplan generation via a diffusion model with discrete and continuous denoising. *arXiv preprint arXiv:2211.13287* (2022). 2
- [VSP*17] VASWANI A., SHAZEER N., PARMAR N., USZKOREIT J., JONES L., GOMEZ A. N., KAISER Ł., POLOSUKHIN I.: Attention is all you need. *Advances in neural information processing systems* (2017). 1
- [WDS*20] WOLF T., DEBUT L., SANH V., CHAUMOND J., DELANGUE C., MOI A., CISTAC P., RAULT T., LOUF R., FUNTOWICZ M., ET AL.: Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations* (2020), pp. 38–45. 2
- [WFT*19] WU W., FU X.-M., TANG R., WANG Y., QI Y.-H., LIU L.: Data-driven interior plan generation for residential buildings. *ACM Transactions on Graphics (TOG)* 38, 6 (2019), 1–12. 1, 3
- [WYN20] WANG X., YESHWANTH C., NIESSNER M.: SceneFormer: Indoor Scene Generation with Transformers. *CoRR abs/2012.09793* (2020). arXiv:2012.09793. 2