



Optimizing for Orientation in Complex Spaces

Xuexin Duan^(✉) and Patrik Schumacher

Tongji University/Zaha Hadid Architects, Shanghai, China
duanxuexin@gmail.com

Abstract. In response to the increasing demand for collaboration and knowledge exchange within Postfordist network society, both virtual and physical spaces are becoming more and more complex. Therefore the orientation within these increasingly complex and information-rich scenes becomes a problem that architectural design must address. The goal of this research is to upgrade architectural design competency in this respect by setting up a workflow for evaluating and optimizing the legibility of complex scenes. This paper introduces a novel research approach focused on the recognizability of salient interaction offerings within complex spatial settings, by using machine learning. A systematic workflow is being developed for simulations that appraise and rank design proposals with respect to the trade-off between scene complexity and legibility. The authors explore the research through a series of simulation experiments concerned with semantic segmentation, i.e. with distinguishing and classifying relevant features in a large complex visual field. The paper first describes the method of setting up the measurement of complexity and ease of recognition, and then illustrates how a trained neural network can be used to evaluate and rank a series of design proposals (with systematically varied degree of complexity) on the basis of their recognizability. While the paper found that the hypothesis of a statistical inverse correlation or trade-off between complexity and recognizability holds, for each degree of complexity there are several design options with different degrees of recognizability. Therefore this approach allows to optimize the design of complex scenes in terms of the crucial criterion of legibility.

Keywords: Orientation · Legibility · Complexity · Spatial cognition · Machine learning · Design optimization

1 Introduction

In response to the increasing demand for collaboration and knowledge exchange within Post Fordist network society, spaces are becoming more and more complex. Therefore the orientation within these increasingly complex and information-rich scenes becomes a problem that architectural design must address. This poses a serious challenge to human spatial perception. Research in spatial cognition, based on observation, involves a variety of behavioural measures. Beyond mere observation and performance measurement spatial cognition theory puts forward hypotheses about how people construct internal representations of space. [5, 12, 13]. The construction of internal representations becomes increasingly challenging as scene complexity increases.

© The Author(s) 2024

C. Yan et al. (Eds.): CDRF 2023, *Phygital Intelligence*, pp. 153–162, 2024.

https://doi.org/10.1007/978-981-99-8405-3_13

As Friedman et al. [6] have argued, it is important to consider the legibility of an environment as a potentially important criterion when evaluating a post-occupancy environment. Many studies have shown that the complexity of a building plan is one of the most important factors affecting orientation in space [18, 15, 19]. However, people's spatial perceptions sometimes do not match the results from the analysis of the 2D plan. Prior research, e.g. Weisman [18] and Passini [16], had limited the study of the relationship between 'plan configuration' and navigation to corridor-type spaces. Their approach is difficult to implement in a more three-dimensional space, such as a complex atrium space. This paper extends the study of the relationship between legibility and architectural plan for atrium spaces. Our investigation of legibility is based on the first person perspective of situated human perception.

Spatial understanding can be partitioned into discrete components instead of being integrated into a single map [11, 2]. One of the components is to locate and orient yourself in a space, which requires recognition of specific spatial features. In recent years, Deep Learning-based Convolutional Neural Networks have shown that trained models are able to recognize objects with human-level accuracy and speed [9, 10]. Recent research in deep-learning based semantic segmentation systems has focused on autonomous driving, industrial inspection, and medical imaging analysis [14, 7]. Architectural design has also drawn considerable attention to deep-learning-based semantic segmentation methods. For example, classifying furniture pictures based on design features [8], and comparing the visual similarity of interior design elements [4]. However, research in this area has focused on demonstrating the capability of successfully recognizing designs in the interior design field. And this study shows the potential of utilizing this technique to recognize salient interaction offerings within complex spatial settings.

The authors explore the research through a series of simulation experiments. We start with a method for measuring complexity. Then we present a method to appraise ease of orientation, and illustrate how a trained neural network can be used to do this, and thereby allowing us to evaluate and rank a series of design proposals.

2 Methodology

The goal of this research is to upgrade architectural design competency by setting up a workflow for evaluating and optimizing the legibility of complex scenes. First the authors set up the formula for calculating the variety value which represents the diversity (complexity) of the destination in space. Then, using a deep-learning-based approach, the authors set up the evaluation model for ease of orientation, which appraises the level of legibility of a given space. The orientation evaluation includes three sub-measurement values: visibility, learnability, and recognizability.

2.1 The Quantitative Definition of Variety

Various research studies have demonstrated that the complexity of the floor plan influences the ease with which users can navigate within a building [15, 18, 16]. This paper interprets complexity as diversity of spatial situations. In retail space, variety means

many different kinds of units instead of mere repetition. Based on the K-means clustering algorithm, the measurement algorithm extracts four geometric attributes from the unit floor plan as input data: area, perimeter, shape proxy “width”, shape proxy “length”. Afterwards, these 4 numbers are combined to create clusters. The algorithm outputs how many types of the units can be clustered. Lastly, it will normalize the number of types as a value between 0 and 1. For instance, a floor plan with 10 units where all units are the same will result in 1 unit type, and variety value 0, i.e. this space is very simple. And when clustered into ten types, its variety value will increase to 1, which indicates a very complex space. Based on this we measure the degree of complexity of a space.

2.2 The Evaluation of Ease of Orientation

Our comparative appraisal of legibility or ease of orientation includes three aspects, namely visibility (a mere geometric precondition), learnability, and recognizability.

2.2.1 Visibility

For defining the visibility value, the authors utilize the Unity Perception package and the Unity Dataset Insights package. The workflow is as follows: First, we systematically build up an array of complex scenes. We then test/compare the scenes as follows: Step 1, place random targets. Targets represent attractions in the space, such as shops, kiosks, etc. Step 2: Place the camera viewpoints randomly within the space at eye level, so that all possible human perceptions can be sampled. Step 3: Assign a color code to all the target objects. As the simulation runs, each frame generates one view image and one labeled image. The data includes labeling information, color coding, and pixel numbers. Second, load the data and run the script. It calculates the total pixels of all the targets automatically using a Python script. So the higher the total pixel count of the targets, the more targets will appear in people’s cone of vision, which means here the higher the space’s visibility value. Third, normalize the data to yield a value between 0 and 1 that represents the comparative level of visibility. This is a simple, objective measure.

2.2.2 Learnability

As mentioned above, deep learning as part of machine learning algorithms has made significant progress in recent years in improving the accuracy of image semantic segmentation. In this research, the authors developed a new estimation method by constructing and training a deep learning network for semantic segmentation of architectural interior images. The authors use this first to derive a learnability value that simulates how fast a human can recognize the features of a new environment.

The measurement process includes two main parts. First, training data set generation. A collection of images and a collection of pixel-labeled images are required to train a semantic segmentation network. Instead of manually segmenting each image, we developed a custom script that uses the Unity Perception Package to automate the process. After labeling and attaching specific color coding to the main architectural elements in the 3D scene: the escalator in magenta color, the balustrade in cyan, the glazing in dark blue, etc., we can start running the simulation. In the simulation, viewpoints

are generated randomly in the scene to generate different view images and semantic segmentation images associated with the view. Second, training the network. This step involves loading the view images and semantic segmentations from the previous step to the learning machine, and the goal is to find out how much time the machine needs to learn to achieve a certain accuracy in its recognition/classification of features (Figs. 1 and 2).

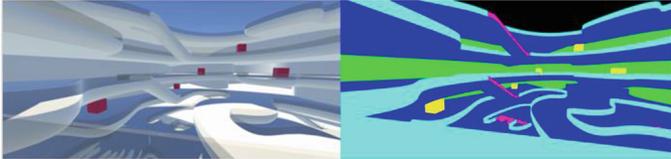


Fig. 1. View image and Semantic Segmentation image of the view.

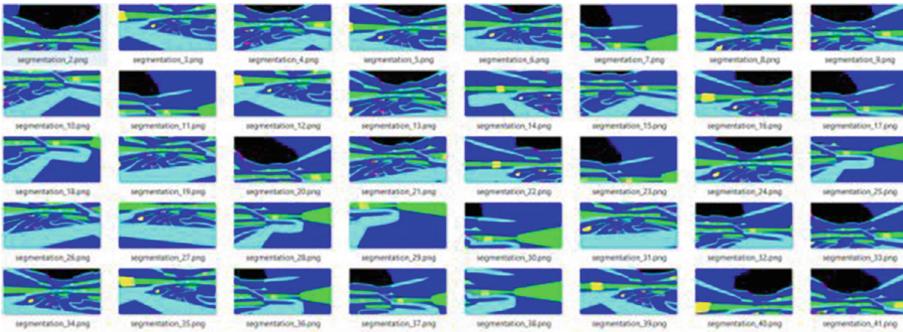


Fig. 2. Semantic segmentation image output for different views

2.2.3 Recognizability

Here, in contrast to comparing how fast a machine can learn and improve to discriminate the architectural elements in different designed spaces, recognizability appraises how well a computer can recognize the architectural elements after it has completed its learning cycle. As previously stated, deep convolutional neural networks are the most powerful method for training a neural network model to extract relevant parts of an image's features. However, how to combine techniques with specific domain knowledge and how to collect data has become a critical issue. To compare how well a trained machine system can recognize specified features in different designed interior spaces, the commonly used deep learning algorithm YOLO (You Only Look Once) was used because of its fast speed and accuracy [17].

In this research, we focus on the escalators as the most significant elements of interior atrium spaces. In our approach, we trained the YOLO model to detect a new custom object (the escalator) based on the trained COCO dataset (Common Objects in Context is a dataset provided by the Microsoft team that can be used for image recognition) instead

of starting from scratch. Prior to deep learning training, data annotation is required. By using the unity perception package, the authors can automatically crop the region and label each image to generate a large amount of training data quickly, rather than manually labeling each image in a common way. There are 600 images prepared, 80% of the total collected data is used for training, and 20% is used for validation.

Instead of using training time data as in the learnability simulation, the detection model solves for identifying (recognizing) the escalators in the view image and in addition outputs a confidence value for each identified element as to how confident the network is in classifying the element as an escalator. The average of all the confidence values generated in a scene feeds into the overall performance measure for this scene, indicating the scene's recognisability (Fig. 3).

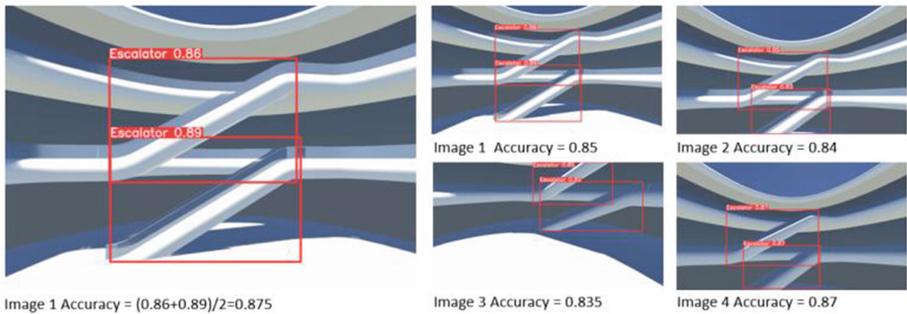


Fig. 3. Example results from detection model, in image 1 the machine detected two escalators with confidence value of 0.86 and 0.89, the accuracy value for this view image 1 is 0.875.

3 Prototype Development and Trade-Off Graph

3.1 Atrium Generator Prototype Setup

The retail atrium was chosen as spatial type capable of exemplifying the problem to be investigated, namely to maintain legibility in the face of complexity. We developed a parametric atrium generator to create a set of systematically differentiated sample spaces. The atrium designs vary the void shape and the depth of the slabs up to the glazing line. The atria are radially divided into 12 segments. In the simplest case these 12 segments are equal. In the most complex case they are all unique.

The value of variety increases with the total number of different segment types. At first the atria operated across two floors above ground, later three floors. A sequence of different design iterations was generated. This sequence measured variety values ranging from 0 to 0.91. In principle, there are unlimited numbers of ways to generate designs with the same variety value. The authors selected three design iterations for each variety value. In total there are 34 design iterations with 12 different variety values (for variety = 0, there is only one design iteration as all the units are the same). For example, in Fig. 4, the variety value increases from left to right, which means the space is becoming more complex in terms of the diversity criteria.



Fig. 4. Plan diagram spectrum for one floor showing the how the variety value increases while the spatial scene becomes more complex.

3.2 Trade-Off Graph

The authors hypothesized that increasing the complexity will reduce the legibility of the space. Indeed, we observed this inverse relationship between complexity and legibility. We see this as a trade-off, since both complexity (delivering diversity of destinations) and legibility (affording ease of orientation) are desirable.

After generating the design iterations with a wide spectrum of variety values, we conducted our learnability and recognizability tests for each of the designs. By mapping all the values for variety first against the obtained learnability and recognizability scores, we get the maps in Fig. 6. Based on statistical theory, when the R^2 value is bigger than 0.6, this means the two parameters have a strong correlation. We also can see from the mappings that for each of the variety values, despite the overall correlation holding, there are several designs with different legibility (learnability/recognizability) scores, i.e. for each visibility value there are relatively better performing designs we might select, i.e. a design with both high complexity and relatively high legibility.

3.2.1 Learnability Trade-off Graph

With the same 34 design iterations spread over 12 design variation values, we executed the learnability experiment to measure how long it takes for the machine to achieve 95% accuracy. For instance, when the variety value = 0. It only took 45s for the machine to learn to recognize the architectural elements and reach an accuracy of 0.95. In comparison, a design with high variety (variety = 0.91), requires 134s of learning to reach the accuracy of 0.95. After training for all the design iterations, we obtained the graph in Fig. 5a which also shows that there is a strong correlation between variety(X axis) and learnability (Y axis). The R^2 value is 0.7461, i.e. significantly above 0.6.

3.2.2 Recognizability Trade-off Graph

Instead of comparing the time taken to reach a certain accuracy between different design iterations, the recognition experiment uses the same trained neural network to test the recognition accuracy with different design iterations. For the same series of designs, 50 images were chosen from each of the 12 design variety brackets. In total, there were 600 images constituting the training dataset. After training, the neural network can achieve an accuracy above 0.8 for the validation set.

After building the neural network, the next step was to test the recognition performance of the neural network for the different design options with different degrees of variety. To test the different designs, 10 new perspectival view images for each design have been fed to the neural network. We then obtained the average accuracy figure for each design, which for us represents the recognizability value of this design. For instance, when the design variety equals 0.00, the average accuracy is 0.84 for the 10 images, and when the design variety equals 0.91, the average accuracy dropped to 0.66. After we run the recognition test for all the options, we get the graph shown in Fig. 5b. We can see from the output figures that the distribution of accuracy correlates with the variety value. This indicates that the recognizability value has a strong correlation with design complexity.

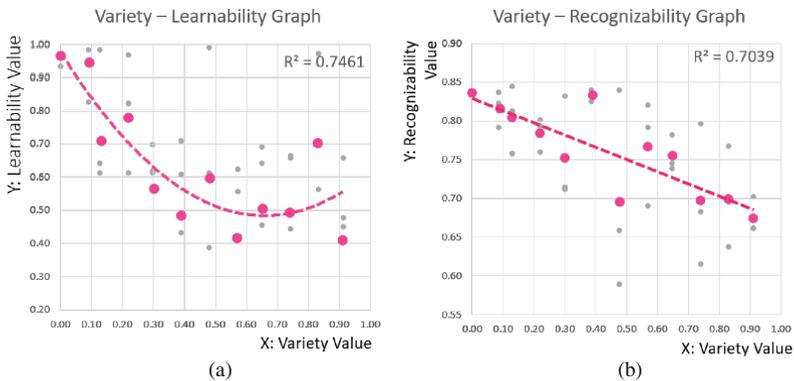


Fig. 5. Variety-Learnability (a), Variety-Recognizability (b) trade-off graphs

Both learnability and recognizability were explored via neural networks trained to semantically segment images of a complex scene and classify features like escalators. According to these experiments, there is a strong inverse correlation between variety and legibility, appraised in its aspects of learnability and recognizability. Since for each variety measure we can generate and test multiple designs, we can use this process to optimize the trade-off between variety and legibility.

4 Further Design Iterations

4.1 Learnability Experiment

It has already been mentioned that there are an infinite number of ways to generate a design for a particular variety value. Our goal is to evaluate and optimize in order to find the most legible solution. The authors compared three design options with similar plan configurations and the same variety value of 0.62, but with changing the slab edge curvature. Option A has a large curvature on the slab edge, Option B has a faceted slab edge, and Option C has a gentle curve on the slab edge. According to the learnability tests, option C exhibits a high learnability value. This is in line with our intuitive expectation

that the smooth version will perform better than the faceted version (because its generates less visual noise).

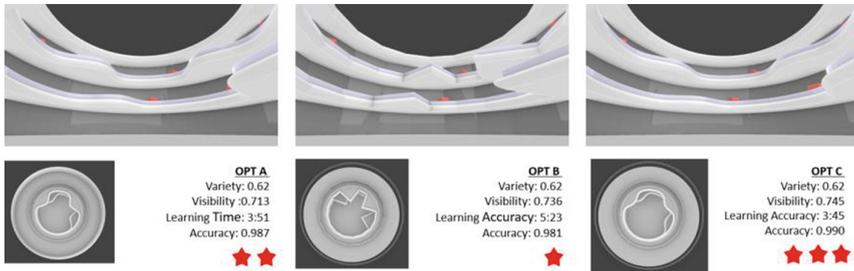


Fig. 6. Three design iterations' legibility comparison. In option C, the visibility value is 0.745, while the other two are 0.713 and 0.736, and in option C, the training time is 3:45 while Option B's training time is much longer as it is 5:23, and Option A is fairly close as it is 3:54.

4.2 Recognizability Experiments

In addition to the prototype experiments, the authors also examined the methods on a real design project. It is a 5-floor atrium space in a mixed-use (retail and cultural) complex. Using two design iterations as examples, we applied our recognisability appraisal method. When all the necessary functional program plan layouts have been met, it is typically depending on the architect's intuitive qualitative appraisal to choose a number of design option that meet the programme criteria. In contrast to this subjective choice, our method allows for reproducible selection based on a quantitative process that can be critiqued and improved upon.

The evaluation and optimization process was applied as follows: first, we randomly selected 10 viewpoints to get 10 view images from each of the designs. Then we input these images to the neural network (trained on the proto-types) to get the escalator recognition performance with accuracy values for each image. Then we averaged all the accuracy values to get an overall recognizability value for the design option in question. Any mistaken detection is taken into account with its confidence value as negative value. Due to mistaken identifications both options' recognizability values were rather low, but this measured difference is significant. For option 1 it was 0.277 and for option 2 it was 0.023 (see Fig. 7). It was interesting that the network was performing at all in a space that was quite different from the training set. In any event, absolute values matter little. The purpose of our method is to get a ranking that guides selection. The result suggests option 1 is better in terms of recognizability. From our intuitive point of view, option 1 also looks more legible. This illustrates how a neural network can be used to evaluate and rank a series of design proposals based on their recognizability.

The evaluation process is quite efficient and easy to operate. You can upload either a screenshot or a rendering of the design space. As a result, a value of recognisability will be output for this space. Many questions and problems remain to be addressed. For instance, when feeding new images that substantially deviate from the original training

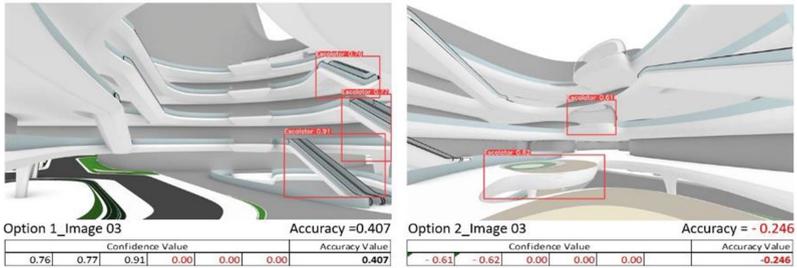


Fig. 7. in image 03 of option 1, the confidence value outputs are 0.76, 0.77, 0.91, but it missed some escalators, so the accuracy value is 0.407 ($= (0.76 + 0.77 + 0.91 + 0 + 0 + 0)/6$). And for option 2, it missed out on all the escalators, and also incorrectly “recognized” some balconies as escalators, so the accuracy value for this view image of option2 is -0.246 ($= (-0.61 - 0.62 + 0 + 0 + 0)/5$).

set, overall accuracy is not high. Hence, it will be necessary to expand the database with more interior images of atriums in the future in order to enhance the accuracy of the data. Second, the current neural network is only trained to recognize the escalator as a key element of the atrium. However, it is possible to extend it to incorporate other key architectural elements as well (Fig. 8).



Fig. 8. Example outputs from detection model tested on the photorealistic renderings.

5 Conclusion and Future Work

The objective of this research is to provide a methodical way of posing the issue of legibility. Our work so far only maps out an initial sketch schematic for the operationalisation of the legibility criterion in its relation to complex scenes. The atrium typology is just one scenario among many. This method setup can be abstracted and generalized, and then tailored to support all urban, architectural and interior design processes. Both complexity (spatial diversity) and legibility (ease of orientation) are considered positive characteristics. While the hypothesis of an inverse correlation or trade-off between complexity and legibility holds, there are several design options with different degrees of legibility for each level of complexity. The methodology can thus help find the sweet spot in the trade-off between these two requirements, or else we can take a certain level of complexity as given, and then identify design options that maximise legibility without

compromising complexity. In this way, the design of complex scenes can be optimized according to the crucial criterion of legibility.

References

1. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al.: Tensorflow: Large-scale machine learning on heterogeneous distributed systems. (2016)
2. Appleyard, D.: Styles and methods of structuring a city. *Environ. Behav.* **2**(1), 100–117 (1970)
3. Arthur, P., Passini, R.: Wayfinding: people, signs, and architecture. (1992)
4. Bell, S., Bala, K.: Learning visual similarity for product design with convolutional neural networks. *ACM Trans Graph TOG.* **34**(4), 1–10 (2015)
5. Denis, M.: Space and spatial cognition: A multidisciplinary perspective. Routledge, (2017)
6. Friedmann, A., Zimring, C., Zube, E.H.: Environmental design evaluation. vol. 9. Springer, (1978)
7. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference, pp. 3431–3440. (2015)
8. Kim, J.S., Song, J.Y., Lee, J.K.: Approach to the extraction of design features of interior design elements using image recognition technique: CAADRIA 2018, pp. 287–296. (2018)
9. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. *Commun. ACM* **60**(6), 84–90 (2017)
10. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436–444 (2015)
11. Lynch, K.: The image of the city. MIT-Press, Cambridge (1960)
12. Mark, D.M.: Human spatial cognition. *Hum Factors Geogr Inf Syst.* 51–60, (1993)
13. Mou, W., McNamara, T.P.: Intrinsic frames of reference in spatial memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **28**(1), 162–170 (2002)
14. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: 18th International conference, proceedings, Part III 18, pp. 234–241. Springer, (2015)
15. O’Neill, M.J.: Effects of signage and floor plan configuration on wayfinding accuracy. *Environ. Behav.* **23**(5), 553–574 (1991)
16. Passini, R.: Wayfinding in architecture. (1984)
17. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference, p. 779–788. (2016)
18. Weisman, J.: Evaluating architectural legibility: Way-finding in the built environment. *Environ. Behav.* **13**(2), 189–204 (1981)
19. Werner, S., Long, P.: Cognition meets le Corbusier—Cognitive principles of architectural design. In: Spatial Cognition III, pp. 112–126. Springer Berlin Heidelberg, (2003)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

