

L^AT_EX Author Guidelines for CVPR Proceedings

Anonymous CVPR submission

Paper ID *****

Abstract

The ABSTRACT is to be in fully justified italicized text, at the top of the left-hand column, below the author and affiliation information. Use the word “Abstract” as the title, in 12-point Times, boldface type, centered relative to the column, initially capitalized. The abstract is to be in 10-point, single-spaced type. Leave two blank lines after the Abstract, then begin the main text. Look at previous CVPR abstracts to get a feel for style and length.

1. 引言

引言部分

1.1. Language

All manuscripts must be in English.

1.2. Dual submission

Please refer to the author guidelines on the CVPR 2022 web page for a discussion of the policy on dual submissions.

1.3. Paper length

Papers, excluding the references section, must be no longer than eight pages in length. The references section will not be included in the page count, and there is no limit on the length of the references section. For example, a paper of eight pages with two pages of references would have a total length of 10 pages.

There will be no extra page charges for CVPR 2022.

Overlength papers will simply not be reviewed. This includes papers where the margins and formatting are deemed to have been significantly altered from those laid down by this style guide. Note that this L^AT_EX guide already sets figure captions and references in a smaller font. The reason such papers will not be reviewed is that there is no provision for supervised revisions of manuscripts. The reviewing process cannot determine the suitability of the paper for presentation in eight pages if it is reviewed in eleven.

1.4. The ruler

The L^AT_EX style defines a printed ruler which should be present in the version submitted for review. The ruler is provided in order that reviewers may comment on particular lines in the paper without circumlocution. If you are preparing a document using a non-L^AT_EX document preparation system, please arrange for an equivalent ruler to appear on the final output pages. The presence or absence of the ruler should not change the appearance of any other content on the page. The camera-ready copy should not contain a ruler. (L^AT_EX users may use options of cvpr.sty to switch between different versions.)

Reviewers: note that the ruler measurements do not align well with lines in the paper — this turns out to be very difficult to do well when the paper contains many figures and equations, and, when done, looks ugly. Just use fractional references (*e.g.*, this line is

087.5), although in most cases one would expect that the approximate location will be adequate.

1.5. Paper ID

Make sure that the Paper ID from the submission system is visible in the version submitted for review (replacing the “*****” you see in this document). If you are using the L^AT_EX template, **make sure to update paper ID in the appropriate place in the tex file.**

1.6. Mathematics

Please number all of your sections and displayed equations as in these examples:

$$E = m \cdot c^2 \quad (1)$$

and

$$v = a \cdot t. \quad (2)$$

It is important for readers to be able to refer to any particular equation. Just because you did not refer to it in the text does not mean some future reader might not need to refer to it. It is cumbersome to have to use circumlocutions like “the equation second from the top of page 3 column 1”. (Note that the ruler will not be present in the final copy, so is not an alternative to equation numbers). All authors will benefit from reading Mermin’s description of how to write mathematics: <http://www.pamitc.org/documents/mermin.pdf>.

1.7. Blind review

Many authors misunderstand the concept of anonymizing for blind review. Blind review does not mean that one must remove citations to one’s own work—in fact it is often impossible to review a paper unless the previous citations are known and available.

Blind review means that you do not use the words “my” or “our” when citing previous work. That is all. (But see below for tech reports.)

Saying “this builds on the work of Lucy Smith [1]” does not say that you are Lucy Smith; it says that you are building on her work. If you are Smith and

Jones, do not say “as we show in [7]”, say “as Smith and Jones show in [7]” and at the end of the paper, include reference 7 as you would any other cited work.

An example of a bad paper just asking to be rejected:

An analysis of the frobnicatable foo filter.

In this paper we present a performance analysis of our previous paper [1], and show it to be inferior to all previously known methods. Why the previous paper was accepted without this analysis is beyond me.

[1] Removed for blind review

An example of an acceptable paper:

An analysis of the frobnicatable foo filter.

In this paper we present a performance analysis of the paper of Smith *et al.* [1], and show it to be inferior to all previously known methods. Why the previous paper was accepted without this analysis is beyond me.

[1] Smith, L and Jones, C. “The frobnicatable foo filter, a fundamental contribution to human knowledge”. *Nature* 381(12), 1-213.

If you are making a submission to another conference at the same time, which covers similar or overlapping material, you may need to refer to that submission in order to explain the differences, just as you would if you had previously published related work. In such cases, include the anonymized parallel submission as supplemental material and cite it as

[1] Authors. “The frobnicatable foo filter”, F&G 2014 Submission ID 324, Supplied as supplemental material `fg324.pdf`.

Finally, you may feel you need to tell the reader that more details can be found elsewhere, and refer them to a technical report. For conference submissions, the paper must stand on its own, and not *require* the reviewer to go to a tech report for further details. Thus,

you may say in the body of the paper “further details may be found in ”. Then submit the tech report as supplemental material. Again, you may not assume the reviewers will read this material.

Sometimes your paper is about a problem which you tested using a tool that is widely known to be restricted to a single institution. For example, let’s say it’s 1969, you have solved a key problem on the Apollo lander, and you believe that the CVPR70 audience would like to hear about your solution. The work is a development of your celebrated 1968 paper entitled “Zero-g frobnication: How being the only people in the world with access to the Apollo lander source code makes us a wow at parties”, by Zeus *et al.*

You can handle this paper like any other. Do not write “We show how to improve our previous work [Anonymous, 1968]. This time we tested the algorithm on a lunar lander [name of lander removed for blind review]”. That would be silly, and would immediately identify the authors. Instead write the following:

We describe a system for zero-g frobnication. This system is new because it handles the following cases: A, B. Previous systems [Zeus et al. 1968] did not handle case B properly. Ours handles it by including a foo term in the bar integral.

...

The proposed system was integrated with the Apollo lunar lander, and went all the way to the moon, don’t you know. It displayed the following behaviours, which show how well we solved cases A and B: ...

As you can see, the above text follows standard scientific convention, reads better than the first version, and does not explicitly name you as the authors. A reviewer might think it likely that the new paper was written by Zeus *et al.*, but cannot make any decision based on that guess. He or she would have to be sure that no other authors could have been contracted to solve problem B.

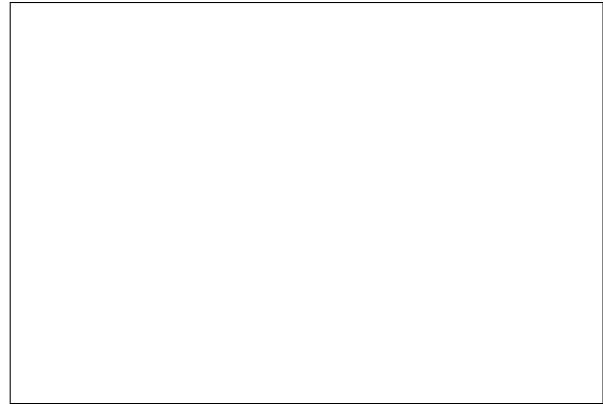


图 1. Example of caption. It is set in Roman so that mathematics (always set in Roman: $B \sin A = A \sin B$) may be included without an ugly clash.

FAQ

Q: Are acknowledgements OK?

A: No. Leave them for the final copy.

Q: How do I cite my results reported in open challenges? **A:** To conform with the double-blind review policy, you can report results of other challenge participants together with your results in your paper. For your results, however, you should not identify yourself and should not mention your participation in the challenge. Instead present your results referring to the method proposed in your paper and draw conclusions based on the experimental comparison to other results.

1.8. Miscellaneous

Compare the following:

`conf_a` conf_a

`$\mathit{conf_a}$` conf_a

See The \TeX book, p165.

The space after *e.g.*, meaning “for example”, should not be a sentence-ending space. So *e.g.* is correct, *e.g.* is not. The provided `\eg` macro takes care of this.

When citing a multi-author paper, you may save space by using “et alia”, shortened to “*et al.*” (not “*et. al.*” as “*et*” is a complete word). If you use the `\etal`

macro provided, then you need not worry about double periods when used at the end of a sentence as in Alpher *et al.* However, use it only when there are three or more authors. Thus, the following is correct: “Frobnication has been trendy lately. It was introduced by Alpher, and subsequently developed by Alpher and Fotheringham-Smythe, and Alpher *et al.* [?].”

This is incorrect: “... subsequently developed by Alpher *et al.* [?] ...” because reference [?] has just two authors.

2. 相关工作

2.1. Margins and page numbering

All printed material, including text, illustrations, and charts, must be kept within a print area $6\frac{7}{8}$ inches (17.46 cm) wide by $8\frac{7}{8}$ inches (22.54 cm) high. Page numbers should be in the footer, centered and $\frac{3}{4}$ inches from the bottom of the page. The review version should have page numbers, yet the final version submitted as camera ready should not show any page numbers. The L^AT_EX template takes care of this when used properly.

2.2. Type style and fonts

Wherever Times is specified, Times Roman may also be used. If neither is available on your word processor, please use the font closest in appearance to Times to which you have access.

MAIN TITLE. Center the title $1\frac{3}{8}$ inches (3.49 cm) from the top edge of the first page. The title should be in Times 14-point, boldface type. Capitalize the first letter of nouns, pronouns, verbs, adjectives, and adverbs; do not capitalize articles, coordinate conjunctions, or prepositions (unless the title begins with such a word). Leave two blank lines after the title.

AUTHOR NAME(s) and AFFILIATION(s) are to be centered beneath the title and printed in Times 12-point, non-boldface type. This information is to be followed by two blank lines.

The **ABSTRACT** and **MAIN TEXT** are to be in a two-column format.

MAIN TEXT. Type main text in 10-point Times, single-spaced. Do NOT use double-spacing. All paragraphs should be indented 1 pica (approx. $\frac{1}{6}$ inch or 0.422 cm). Make sure your text is fully justified—that is, flush left and flush right. Please do not place any additional blank lines between paragraphs.

Figure and table captions should be 9-point Roman type as in Figs. 1 and 2. Short captions should be centred.

Callouts should be 9-point Helvetica, non-boldface type. Initially capitalize only the first word of section titles and first-, second-, and third-order headings.

FIRST-ORDER HEADINGS. (For example, **1. Introduction**) should be Times 12-point boldface, initially capitalized, flush left, with one blank line before, and one blank line after.

SECOND-ORDER HEADINGS. (For example, **1.1. Database elements**) should be Times 11-point boldface, initially capitalized, flush left, with one blank line before, and one after. If you require a third-order heading (we discourage it), use 10-point Times, boldface, initially capitalized, flush left, preceded by one blank line, followed by a period and your text on the same line.

2.3. Footnotes

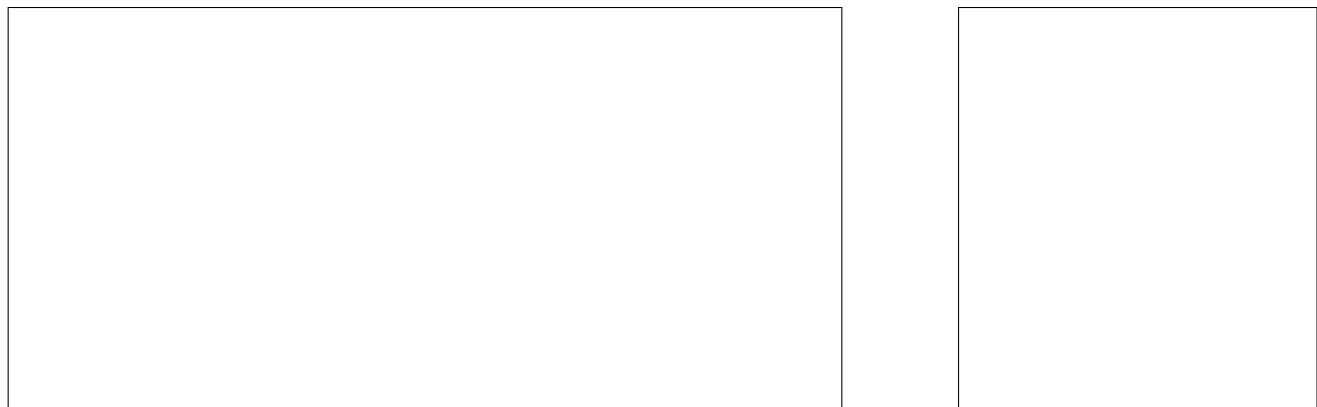
Please use footnotes¹ sparingly. Indeed, try to avoid footnotes altogether and include necessary peripheral observations in the text (within parentheses, if you prefer, as in this sentence). If you wish to use a footnote, place it at the bottom of the column on the page on which it is referenced. Use Times 8-point type, single-spaced.

2.4. Cross-references

For the benefit of author(s) and readers, please use the

`\cref{...}`

¹This is what a footnote looks like. It often distracts the reader from the main flow of the argument.



(a) An example of a subfigure.

(b) Another example of a subfigure.

图 2. Example of a short caption, which should be centered.

Method	Frobnability
Theirs	Frumpy
Yours	Frobbly
Ours	Makes one's heart Frob

表 1. Results. Ours is better.

command for cross-referencing to figures, tables, equations, or sections. This will automatically insert the appropriate label alongside the cross-reference as in this example:

To see how our method outperforms previous work, please see Fig. 1 and Tab. 1. It is also possible to refer to multiple targets as once, *e.g.* to Figs. 1 and 2a. You may also return to Sec. 2 or look at Eq. (2).

If you do not wish to abbreviate the label, for example at the beginning of the sentence, you can use the

`\Cref{...}`

command. Here is an example:

Figure 1 is also quite important.

2.5. Illustrations, graphs, and photographs

All graphics should be centered. In \LaTeX , avoid using the `center` environment for this purpose, as this adds potentially unwanted whitespace. Instead use

`\centering`

at the beginning of your figure. Please ensure that any point you wish to make is resolvable in a printed copy of the paper. Resize fonts in figures to match the font in the body text, and choose line widths that render effectively in print. Readers (and reviewers), even of an electronic copy, may choose to print your paper in order to read it. You cannot insist that they do otherwise, and therefore must not assume that they can zoom in to see tiny details on a graphic.

When placing figures in \LaTeX , it's almost always best to use `\includegraphics`, and to specify the figure width as a multiple of the line width as in the example below

```
\usepackage{graphicx} ...
\includegraphics[width=0.8\linewidth]
{myfile.pdf}
```

2.6. Color

Please refer to the author guidelines on the CVPR 2022 web page for a discussion of the use of color in your document.

If you use color in your plots, please keep in mind that a significant subset of reviewers and readers may have a color vision deficiency; red-green blindness is the most frequent kind. Hence avoid relying only on color as the discriminative feature in plots (such as red *vs.*

green lines), but add a second discriminative feature to ease disambiguation.

3. 方法论

3.1. 建筑物立面语义信息提取

传统的语义分割方法通过像素级的分类来分配标签, 能够为图像中的每个像素点提供详细的语义信息。现有的立面语义分割的研究中, 通常先对立面进行语义分割, 然后分析分割结果的连通语义区域, 对每个连通区域计算 bounding box 框定语义部分 [5,6,9]。这种方法能够提供较为精准的像素级语义标签, 然而, 由于分割结果会产生零碎且不连续的语义块, 对于每块语义区域计算的边界框也是零碎离散的, 无法保证构件级别的语义精准度, 引入不必要的误差和噪声。因此我们采用 YOLO 架构 [8] 目标检测方法提取立面语义, 避免分割方法中的零碎问题。

3.2. 建筑物立面语义信息规则化

3.2.1 立面 bounding box 分组

在提取立面语义信息后, 对得到的识别框进行立面规则化, 使识别框排列分布符合一致性。首先需要对面元素进行分组, 分组过程中需要保证分组内的元素尽量一致。我们将检测框的长宽信息映射到二维平面上, 形成一个二维特征空间, 采用 meanshift 聚类方法 [2] 对检测框进行分类, 并将 meanshift 邻域窗口采用矩形框的形式, 尺寸设置为当前组中的平均长、宽的 λ 倍, 这样能够确保采用 meanshift 方法划分每个组中立面元素具有一致的长宽比。

$$\bar{L} = \frac{1}{N} \sum_{i=1}^N L_i \quad (3)$$

$$\bar{W} = \frac{1}{N} \sum_{i=1}^N W_i \quad (4)$$

$$\text{Window Size} = \lambda \left(\frac{1}{N} \sum_{i=1}^N W_i, \frac{1}{N} \sum_{i=1}^N L_i \right) \quad (5)$$

$$m(x_i) = \frac{\sum_{x_j \in W(x_i)} K(x_i - x_j) x_j}{\sum_{x_j \in W(x_i)} K(x_i - x_j)} \quad (6)$$

参考 [9] 立面相似元素的做法, 我们将 λ 设为五分之一。这样就将不同尺度的窗户给区分开。此外, 根

据格式塔理论的邻近性和相似性原则, 相似的立面构件元素在水平和垂直方向上具有一致的长或宽属性 [6], 因而需要进一步将相似元素在水平和垂直方向上分组, 以通过能量函数对齐相似元素。分组方式与之前的方法相似, 若根据公式 6 划分出组 $E: \{e_1, e_2, e_3, \dots, e_n\}$, 计算每个元素的质心坐标并投影到二维平面上, 同样采各分组阈值 Window Size 来划分水平分组 g_x 和垂直分组 g_y 。

3.2.2 分组间 bounding box 优化

组间优化能量项出处 [1] 基于的启发, 分组间的一致性可以用分组间的最佳位移交并比 IOU 来衡量。

$$R_{ij} = \frac{R(S_i \cap T(S_j))}{R(S_i \cup T(S_j))} \quad (7)$$

其中, R_{ij} 表示分组 S_i 与 S_j 经过最佳变换 T 后的交并比。能量函数可定义为:

3.2.3 组内 bounding box 优化

通过 3.2.1 得到水平分组 g_x 和垂直分组 g_y 后, 定义能量函数优化各分组的对齐性。根据水平分组 g_x 和垂直分组 g_y 的对齐性, 边界框的水平分组中应具有相同的水平起始位置和宽度, 垂直分组中应具有相同的垂直起始位置和高度。我们采用牛顿法迭代优化方法进行立面规则优化 [3]。

$$E_x = \sum_{g_i \in G} \sum_{i \in g_i} \left[(x_i - \bar{x})^2 + (w_i - \bar{w})^2 \right] \quad (8)$$

$$E_y = \sum_{g_i \in G} \sum_{i \in g_i} \left[(y_i - \bar{y})^2 + (h_i - \bar{h})^2 \right] \quad (9)$$

$$E_{total} = E_x + E_y \quad (10)$$

式中, $\bar{x}, \bar{w}, \bar{y}, \bar{h}$ 分别表示该水平或者垂直分组中的水平方向横坐标均值, 宽度均值, 左上角纵坐标均值以及高度均值。最后总的能量项 E_{total} 为水平能量项 E_x 和垂直能量项 E_y 之和。

3.3. 窗户语义分割方法

在窗户语义分割任务中, 立面窗户玻璃与语义属于二分类问题, 较为简单。因而采用图像分割的方法提取窗户区域的语义信息, 包括窗框和玻璃。传统的图像

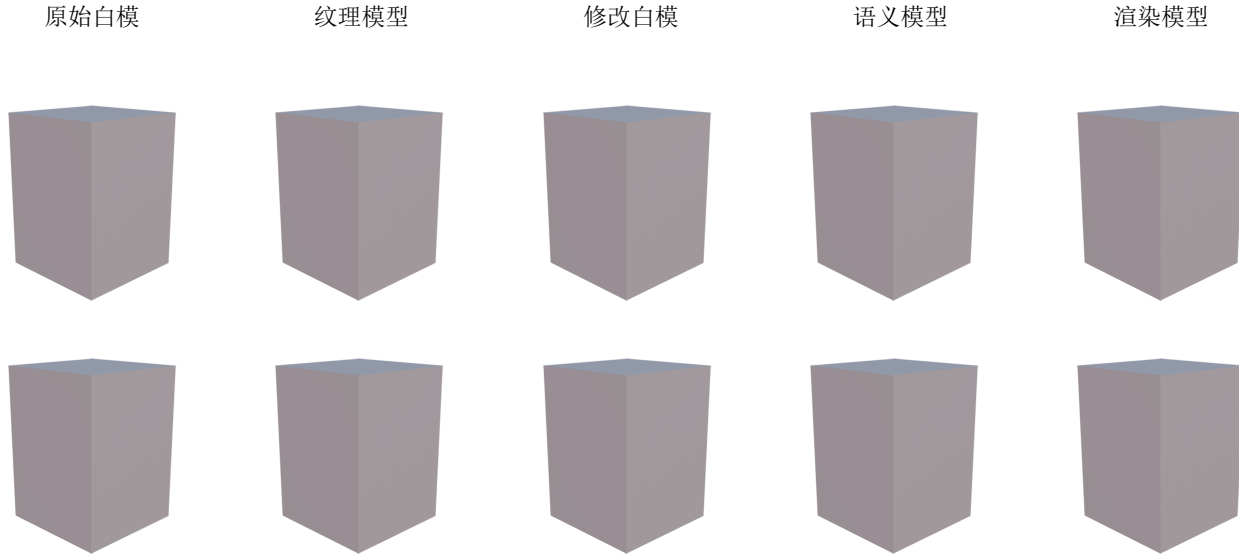


图 3. 语义建模效果示意图

分割方法包括区域生长、阈值法、分水岭算法、边缘检测算法，由于窗户影像颜色梯度变换较多，边缘检测算法和分水岭算法在生成过程中会产生较多的毛躁点，无法有效区分窗框及玻璃信息，而区域生长算法难以用单一的度量指标适用不同场景。我们观察到，在大量的窗户影像中通常有着明暗区分的区域，因此采用基于 otsu 阈值法来进行窗户语义提取 [7]。使用 otsu 阈值法并结合形态学膨胀操作能够基本提取窗户语义信息，但仍存在微小噪点。为了进一步提取干净的窗户与玻璃语义，我们采用金字塔纹理平滑 [11] 的方法对窗户影像做预处理，再通过 otsu 阈值法进行二值化分割，并利用形态学开运算操作去除二值图中的微小噪声，最终得到窗户的分割掩码。随后，采用连通性分析计算每个连通区域，以 bounding box 作为每个连通区域的玻璃实例，并采用参数表示 $\{x, y, w, h\}$ 。

3.4. 建筑构件空间信息生成

我们采用预训练的 Metric3D 模型 [4,10] 估计立面图像深度信息与法线信息，并结合 YOLO 的预测框计算为每个类计算平均深度信息，以立面深度为基准，转换窗户、门、玻璃等类别的相对深度信息。

此外，玻璃的位置信息是相对于当前窗户实例的相对位置，为了统一各个参数信息为相对于立面参数，需要进行简单转换。

对于单个建筑模型，我们构建了层次化的参数化数

参数	描述
立面层	tw 纹理包围盒区域的宽度
	th 纹理包围盒区域的高度
	pw 立面多边形的实际宽度
	ph 立面多边形的实际高度
类别层	n 立面的法线
	C 类别
	Depth 类别深度
实例层	x box 左上角横坐标
	y box 左上角纵坐标
	w box 宽度
	h box 高度

表 2. 立面层次化的参数结构

据结构，以表示模型的立面构件的空间位置信息，参见表 2。 <https://www.computer.org/about/contact>.

4. 实验

5. 结论

参考文献

- [1] Jun Cao, Henning Metzmacher, James O'Donnell, Jérôme Frisch, Vladimir Bazjanac, Leif Kobbelt, and

- Christoph van Treeck. Facade geometry generation from low-resolution aerial photographs for building energy modeling. *Building and Environment*, 123:601–624, 2017. 6
- [2] Dorin Comaniciu and Peter Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 24(5):603–619, 2002. 6
- [3] Roger Fletcher. *Practical methods of optimization*. John Wiley & Sons, 2000. 6
- [4] Mu Hu, Wei Yin, Chi Zhang, Zhipeng Cai, Xiaoxiao Long, Hao Chen, Kaixuan Wang, Gang Yu, Chunhua Shen, and Shaojie Shen. Metric3d v2: A versatile monocular geometric foundation model for zero-shot metric depth and surface normal estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 7
- [5] Wenguang Ma, Shibiao Xu, Wei Ma, Xiaopeng Zhang, and Hongbin Zha. Progressive feature learning for facade parsing with occlusions. *IEEE Transactions on Image Processing*, 31:2081–2093, 2022. 6
- [6] Markus Mathias, Andelo Martinović, and Luc Van Gool. Atlas: A three-layered approach to facade parsing. *International Journal of Computer Vision*, 118:22–48, 2016. 6
- [7] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979. 7
- [8] Ao Wang, Hui Chen, Lihao Liu, Kai Chen, Zijia Lin, Jungong Han, et al. Yolov10: Real-time end-to-end object detection. *Advances in Neural Information Processing Systems*, 37:107984–108011, 2025. 6
- [9] Yuefeng Wang, Wei Jiao, Hongchao Fan, and Guoqing Zhou. A framework for fully automated reconstruction of semantic building model at urban-scale using textured lod2 data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 216:90–108, 2024. 6
- [10] Wei Yin, Chi Zhang, Hao Chen, Zhipeng Cai, Gang Yu, Kaixuan Wang, Xiaozhi Chen, and Chunhua Shen. Metric3d: Towards zero-shot metric 3d prediction from a single image. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9043–9053, 2023. 7
- [11] Qing Zhang, Hao Jiang, Yongwei Nie, and Wei-Shi Zheng. Pyramid texture filtering. *ACM Trans. Graph.*, 42(4), July 2023. 7