Writing (Computer Vision) Papers from the Reviewer's Perspective

LatinX in CV Workshop @ ICCV 2021

October 11th, 2021

Dr. Matias Valdenegro

matiasv@latinxinai.org
@mvaldenegro
http://github.com/mvaldenegro

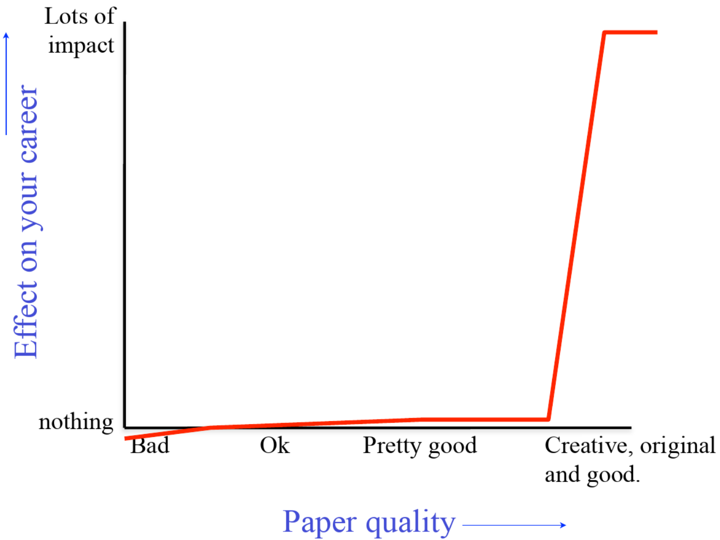
Why do we write papers?

- To advance personal careers (Grad school, PhD, etc)?
- To make an impact in the field?
- To communicate ideas and advance the state of the art?

The main idea writing of scientific papers is to communicate ideas to an audience, or review other people's work.

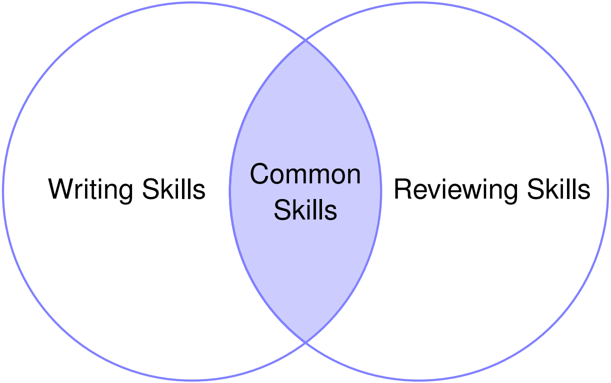
This means that a paper is written not for the author to read, but for other scientists to understand.

Impact of Papers



From "How to write a good CVPR submission" by Bill Freeman, 2014.

Writing and/or Reviewing Papers?



There is a large intersection between both skill sets.

Writing papers requires the reviewer's perspective and reviewing papers requires the author's perspective.

The Seven W's

What is the problem?

Define the problem and its context.

Why is it important?

Describe why the problem is important and to who it is important.

What have other people done about the problem?

Survey the literature for the problem, describing the state of the art in detail. People generally write a conceptual framework to categorize previous research.

The Seven W's

Why is it not sufficient?

Usually part of the introduction or related work, describes why the problem is still not solved, either from a theoretical or practical standpoint. There could be many deficits and people only describe the ones relevant to their paper.

What do you propose to do differently?

This is the main content of the paper, describing the proposed technique, theoretical framework, or research idea and concepts.

Why is it better?

Corresponds to the evaluation section of the paper (experimental or theoretical), and contains the main body of results that argue why the proposed technique is better.

The Seven W's

What is left to be done?

Corresponds to the analysis of results and conclusions/future work sections.

- All these steps/questions should be present in a paper in clear writing.
- It is a good reference for both writing and reviewing a paper.
- After reading the paper, the reviewer should be able to answer these questions to a degree, anything missing might point to an issue in the paper.

The Reviewer's Job

- Check that the paper is correctly evaluated, written, sound methodology, and it is understandable.
- Evaluate and check the claims made in the paper.
- Overall, provide constructive feedback that **improves the paper**.
- Sometimes, evaluate novelty of the proposed approach or gap in the state of the art.
- If there is a rebuttal process, ask questions , interact with the authors, and re-evaluate their review, if needed.

Author(s) vs Reviewer(s)

Author	Reviewer
Drafts, designs, and writes the paper	Provides feedback about the paper.
Knows their own work quite well, worked on it for a long time	"One-shot" look and opinion about the work.
Knows the idea in their head, and tries to put it in words/figures/tables in their paper	Tries to understand the idea behind the paper, just by reading the paper.
Can anticipate the reviewer's requests/thinking	No way to deal with uncertainty in the paper.

Author and reviewer should not be enemies, but friends!

Core Advice

When writing a paper, experienced authors (that usually also act as reviewers), can predict issues the reviewers will point out. This is why writing a paper alone (specially as beginner) can be difficult.

Always have other (more experienced) researchers read your paper and make comments. This very important to get an outsiders perspective.

The very important point is that **other people should understand the paper**, not just the authors. This is the most common issue pointed by reviewers. Audience research/selection is very important.

Methodological Choices

Baselines Selecting appropriate baselines is difficult, and it is a common source of reviewer complaints.

Datasets Use the standard datasets for the task, if deviating, justify appropriately. This can be problematic in new tasks with no available datasets.

Ablations Perform the correct variations of hyper-parameters or algorithmic choices to evaluate your technique/system and find how performance changes and justify your choices.

Making Comparisons

A big source of conflict with reviewers is how comparisons are made, this can be because:

Baselines Incorrect (not SOTA) or improperly tuned baselines can create reviewer uncertainty.

Train/Test Splits Leakage between train/test sets must be prevented, if creating new datasets, then this part is very important. Leakage can happen due to individuals being in both sets, data augmentation, incorrect methodology, etc.

Metrics Each metric being evaluated must be carefully selected and justified, it should produce some knowledge by evaluating it. Using the incorrect metric for a task will be noticed by reviewers.

Aesthetics and Visual Appeal

The visual look and aesthetics of the paper does matter. It will play a role on how the reviewer looks at the paper. Some tips:

- Place figures/tables at the top of the page/column.
- Ensure that figures/images are high resolution and are understandable.
- Format tables without vertical lines, with proper spacing, and explainable rows/columns
- Use colors to guide the reader, for example, by highlighting best performing combinations in **bold**.

Deep Paper Gestalt (Bad Papers)

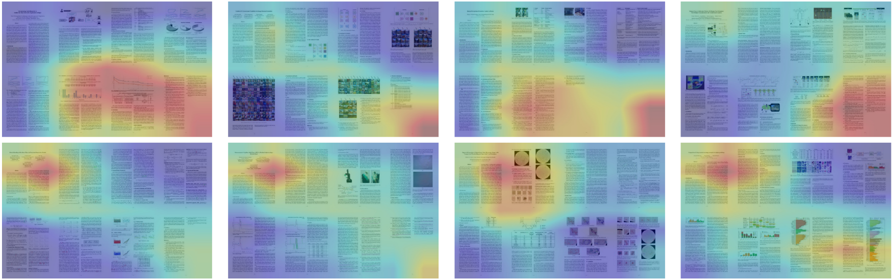


Figure 5. **Class-specific discriminative regions for *bad* papers.** (*Top*) Failing to fill the paper into a full eight-page paper is a discriminative visual cue for bad paper. (*Bottom*) The generated heatmaps focus on the top-right corner of the first page. This suggests that the *absence* of illustrative figures in the first two pages may cause the paper more difficult to understand.

Figure from "Deep Paper Gestalt" by Jia-Bin Huang, arXiv:1812.08775.

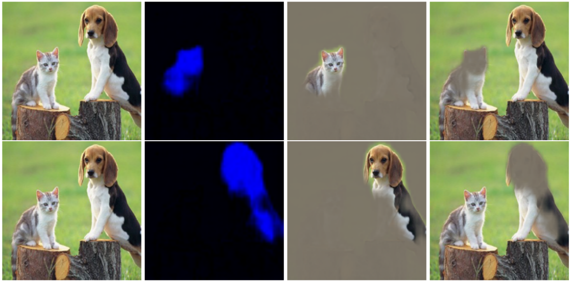
Deep Paper Gestalt (Good Papers)

Figure 6. **Class-specific discriminative regions for *good* papers.** The heatmap generated by class activation mapping [24] highlights regions specific to good papers, e.g., teaser figures in the first page for illustrating the main ideas, tables/plots showing a sense of thoroughness in experimental validation, impressive math equations, and arrays of colorful images for qualitative results from benchmark datasets.

Figure from "Deep Paper Gestalt" by Jia-Bin Huang, arXiv:1812.08775.

"Teaser" Figure One on Page One/Two

Many people put a specially crafted figure/diagram that teases or explains the proposed technique. It helps explain in simple words/diagrams what the paper is about and how it is different from the state of the art. I believe this is a good practice.



(a) Input Image (b) Generated saliency map (c) Image multiplied by the mask (d) Image multiplied by inverted mask

Figure 1: An example of explanations produced by our model. The top row shows the explanation for the "Egyptian cat" while the bottom row shows the explanation for the "Beagle". Note that produced explanations can precisely both highlight and remove the selected object from the image.

Figure from [Dabkowski and Gal. 2017].

Toy Examples

Many papers include examples on small or toy datasets, in order to showcase the basics of the idea/concept. This connects back with the teaser figure, and it gives the reader a small idea of the basic concepts in the paper. Later examples can be more complex.

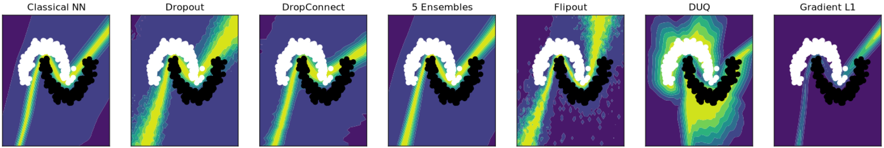


Figure made by myself to show different uncertainty methods on the two moons dataset. Differences are quite clear.

Jia-Bin Huang' Advice

Make appropriate captions that guide the reader about what to "see" in this figure/table.

DON'Ts

Please refer to the texts for detail.

DOs

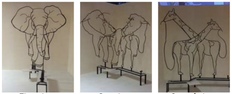


Fig. 2. **Example of multi-view wire sculpture art.** The anamorphose sculpture created by the French sculptor Matthieu Robert-Ortis is a classic example of *multi-view wire art*. When viewing from one specific angle, we perceive a drawing of an elephant. When viewing from another view point, the interpretation changes into two giraffes. The 2D projection in the intermediate view does not produce an interpretable image.

Always include citations in the text to guide the reader to which technique/dataset is being compared.

DON'Ts

We compare our method against X-Net, Y-Net, Z-Net on the dataset W.

DOs

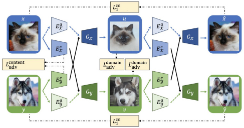
We compare our method against X-Net [1], Y-Net [2], Z-Net [3] on the dataset W [4].

Figures taken from `https://twitter.com/jbhuang0604/status/1279992087497314305`.

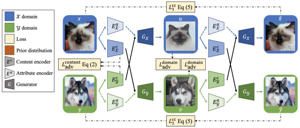
Jia-Bin Huang' Advice

Use a consistent notation and use it in your figures/tables.

DON'Ts

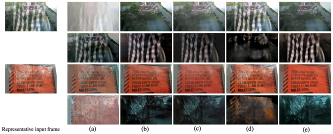


DOs

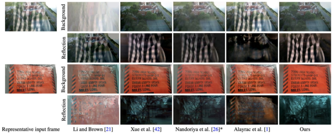


Make it easier for the reader to interpret your figures.

DON'Ts



DOs



Jia-Bin Huang' Advice

Use human-readable notations for equations.

DON'Ts

$$L = \lambda_1 L_1 + \lambda_2 L_2$$
$$\vdots$$
$$\lambda_1 = 0.1, \lambda_2 = 2.5$$

DOs

$$L_{total} = \lambda_{rec} L_{rec} + \lambda_{smooth} L_{smooth}$$
$$\vdots$$
$$\lambda_{rec} = 0.1, \lambda_{smooth} = 2.5$$

Group captions in sub-figures and tables for easier interpretation.

DON'Ts

DOs

Jia-Bin Huang' Advice

Augment math notation in text with human interpretations.

DON'Ts

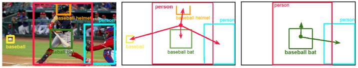
The F_θ takes I_i, I_j as inputs and produce $W_{i\to j}$.

DOs

The flow estimation network F_θ takes a pair of frames I_i, I_j as inputs and produces a dense flow field $W_{i\to j}$.

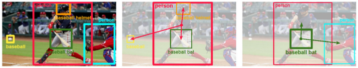
Use image/shape attributes to describe ideas/results.

DON'Ts



(a) Object detection (b) Human-centric (c) Object-centric

DOs



(a) Object detection (b) Human-centric (c) Human-centric

Jia-Bin Huang' Advice

Provide details between differences in the SOTA and your proposed method.

DON'Ts

Method	mAP
[Snorlax et al. 2018]*	25.0
[Bulbasaur et al. 2019]*†	29.8
[Psyduck et al. 2020] †	32.1
Ours	35.5

DOs

Method	External data?	Finetuned?	mAP
[Snorlax et al. 2018]	✓	-	25.0
[Bulbasaur et al. 2019]	✓	✓	29.8
[Psyduck et al. 2020]	-	✓	32.1
Ours	-	-	35.5

One message/ablation per table.

DON'Ts

Method	Acc (%)
Variant A1	75
Variant A2	65
w/o B	77
w/o C	73
Full model	80

DOs

Method	Acc (%)
Full model	80
Variant A1	75
Variant A2	65

Method	Acc (%)
Full model	80
Without B	77

Method	Acc (%)
Full model	80
Without C	73

Jia-Bin Huang' Advice

Format and group table columns in a human understandable way, specially when using related metrics/datasets.

DON'Ts

Method	Abs Rel	Sq Rel	RMSE	log RMSE	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
DeepV2D [56]	0.526	3.629	6.493	0.683	0.487	0.671	0.761
Ours - Single-scale pose (aligned MiDaS)	0.380	2.617	5.773	0.533	0.562	0.736	0.832
Ours - Single-scale pose + depth fine-tuning	0.472	3.444	6.340	0.635	0.534	0.694	0.790
Ours - Single-scale pose + depth filter	**0.375**	**2.546**	**5.763**	**0.530**	**0.569**	0.738	0.835
Ours - Flexible pose	0.379	2.702	5.795	0.533	0.565	0.744	0.836
Ours - Flexible pose + depth fine-tuning	0.439	3.100	6.213	0.614	0.524	0.698	0.796
Ours - Flexible pose + depth filter	0.377	2.657	5.786	0.531	0.568	**0.745**	**0.837**

Method	ATE (m)	RPE Trans (m)	RPE Rot (deg)
DeepV2D [56]	0.9526	0.3819	0.1869
Ours - Single-scale pose (aligned MiDaS)	0.1883	0.0806	0.0262
Ours - Single-scale pose + depth fine-tuning	0.1686	0.0724	0.0139
Ours - Single-scale pose + depth filter	0.1882	0.0806	0.0262
Ours - Flexible pose	0.1843	0.0723	0.0095
Ours - Flexible pose + depth fine-tuning	**0.1656**	**0.0651**	**0.0070**
Ours - Flexible pose + depth filter	0.1843	0.0723	0.0095

DOs

	Depth - Error metric↓				Depth - Accuracy metric↑			Pose - Error metric↓		
Method	Abs Rel	Sq Rel	RMSE	log RMSE	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$	ATE (m)↓	RPE Trans (m)↓	RPE Rot (deg)↓
DeepV2D [56]	0.526	3.629	6.493	0.683	0.487	0.671	0.761	0.9526	0.3819	0.1869
Ours - Single-scale pose (aligned MiDaS)	0.380	2.617	5.773	0.533	0.562	0.736	0.832	0.1883	0.0806	0.0262
Ours - Single-scale pose + depth fine-tuning	0.472	3.444	6.340	0.635	0.534	0.694	0.790	0.1686	0.0724	0.0139
Ours - Single-scale pose + depth filter	**0.375**	**2.546**	**5.763**	**0.530**	**0.569**	0.738	0.835	0.1882	0.0806	0.0262
Ours - Flexible pose	0.379	2.702	5.795	0.533	0.565	0.744	0.836	0.1843	0.0723	0.0095
Ours - Flexible pose + depth fine-tuning	0.439	3.100	6.213	0.614	0.524	0.698	0.796	**0.1656**	**0.0651**	**0.0070**
Ours - Flexible pose + depth filter	0.377	2.657	5.786	0.531	0.568	**0.745**	**0.837**	0.1843	0.0723	0.0095

Ross Girshick's Advice

Ross Girshick gave very good advice at ICCV 2019, South Korea, mostly about Object Detection:

- A paper should be about a single focused idea/question.
- "Idea" usually means method, what should I learn?
  - Under what conditions does it work?
  - When does it **not** work?
  - If the idea has multiple components, which are the most important?
  - Which implementation details are important?
- I seldom care if "your idea + unrelated ideas/tricks" $\rightarrow$ SOTA results.
  - My first priority is to learn some interesting things about your idea.

Reference:

`https://twitter.com/prajjwal_1/status/1188653550810697728`

Ross Girshick's Advice - Simplicity

- Start from a solid baseline.
- Apply your idea to it.
- Perform ablations under simple settings.

This should be the most basic evaluation method for any paper that proposes new methods/techniques.

Ross Girshick's Advice - Ablations

Support all your claims!

- **All claims** should be supported.
  - By citation, or.
  - By experiments (carefully designed).
- Otherwise, qualify the statement:
  - "Intuitively, increasing X is important for Y..."
  - "Increasing X may lead to improved Y..."

Ross Girshick's Advice - Claims

Support all your claims!

- **All claims** should be supported.
  - By citation, or.
  - By experiments (carefully designed).
- Otherwise, qualify the statement:
  - "Intuitively, increasing X is important for Y..."
  - "Increasing X may lead to improved Y..."

Mask R-CNN's Ablation Tables

net-depth-features	AP	AP_{50}	AP_{75}
ResNet-50-C4	30.3	51.2	31.5
ResNet-101-C4	32.7	54.2	34.3
ResNet-50-FPN	33.6	55.2	35.3
ResNet-101-FPN	35.4	57.3	37.5
ResNeXt-101-FPN	**36.7**	**59.5**	**38.9**

(a) **Backbone Architecture**: Better backbones bring expected gains: deeper networks do better, FPN outperforms C4 features, and ResNeXt improves on ResNet.

	AP	AP_{50}	AP_{75}
softmax	24.8	44.1	25.1
sigmoid	**30.3**	**51.2**	**31.5**
	+5.5	*+7.1*	*+6.4*

(b) **Multinomial *vs.* Independent Masks** (ResNet-50-C4): *Decoupling* via per-class binary masks (sigmoid) gives large gains over multinomial masks (softmax).

	align?	bilinear?	agg.	AP	AP_{50}	AP_{75}
RoIPool [9]			max	26.9	48.8	26.4
RoIWarp [7]		✓	max	27.2	49.2	27.1
		✓	ave	27.1	48.9	27.1
RoIAlign	✓	✓	max	**30.2**	**51.0**	**31.8**
	✓	✓	ave	**30.3**	**51.2**	**31.5**

(c) **RoIAlign** (ResNet-50-C4): Mask results with various RoI layers. Our RoIAlign layer improves AP by ~3 points and AP_{75} by ~5 points. Using proper alignment is the only factor that contributes to the large gap between RoI layers.

	AP	AP_{50}	AP_{75}	AP^{bb}	AP^{bb}_{50}	AP^{bb}_{75}
RoIPool	23.6	46.5	21.6	28.2	52.7	26.9
RoIAlign	**30.9**	**51.8**	**32.1**	**34.0**	**55.3**	**36.4**
	+7.3	*+ 5.3*	*+10.5*	*+5.8*	*+2.6*	*+9.5*

(d) **RoIAlign** (ResNet-50-**C5**, *stride 32*): Mask-level and box-level AP using *large-stride* features. Misalignments are more severe than with stride-16 features (Table 2c), resulting in massive accuracy gaps.

	mask branch	AP	AP_{50}	AP_{75}
MLP	fc: 1024→1024→80·28^2	31.5	53.7	32.8
MLP	fc: 1024→1024→1024→80·28^2	31.5	54.0	32.6
FCN	conv: 256→256→256→256→256→80	**33.6**	**55.2**	**35.3**

(e) **Mask Branch** (ResNet-50-FPN): Fully convolutional networks (FCN) *vs.* multi-layer perceptrons (MLP, fully-connected) for mask prediction. FCNs improve results as they take advantage of explicitly encoding spatial layout.

Table 2. **Ablations** for Mask R-CNN. We train on `trainval35k`, test on `minival`, and report *mask* AP unless otherwise noted.

These ablations **justify** the different choices made in this detector.

Reviewer Variability/Uncertainty

In most top conferences, there is significant variability in reviewer opinion, experience, and attention.

- This is due to expert reviewer shortage, good reviewers become area chairs, and finding good reviewers is difficult.
- Also due to the deluge of new papers being submitted to each conference, and increasing every year.
- Very annoying from the author's perspective, as the purpose of the review process is not only to make accept/reject decisions, but also help the author improve their research and paper.
- Only solution possible is for the community to invest more resources (time) into the review process, and to train reviewers.

Research Ideas in CV/ML/AI

- Work/research on interesting topics, find your own niche.
- Consider **societal and research field impact when selecting problems**.
  - If the problem is not important, when solved, only a few people will care.
  - If its a very important problem, even if progress is small, it will have a large importance and impact.
- Consider the impact of AI developments on minorities and disadvantaged groups.
- Think deeply on **who gets power** by an algorithm. Always think and prevent misuses.
- Just improving on the state of the art (a % on a benchmark) is not always the best. **New problems, tasks, and datasets, are very important too!**

Take Home Messages

- Writing papers is an art, and does require a bit of artistic skills to clearly present ideas.
- We write papers so **other people can understand them**, they have a clear audience, and to advance the state of the art.
- Writing, presenting ideas/data/figures, is not as easy as one might think. **Clarity is paramount**.
- It takes time/experience to polish the paper. It is not just about evaluation results, but how they are presented. This is **very** important.
- I usually look at highly cited papers to see how they present results, tables, formatting, figures, etc. This has been time well spent.

Writing/Reviewing Resources

- CVPR 2021 reviewing guidelines.
  http://luthuli.cs.uiuc.edu/~daf/CVPR21TrainingMaterials/RefSlides.pdf
- CVPR 2018 Workshop "Good Citizen of CVPR".
  https://www.cc.gatech.edu/~parikh/citizenofcvpr/
- CVPR 2020 Tutorial on How to Write a Good Review.
  https://sites.google.com/view/making-reviews-great-again/
- How to write a good CVPR submission by Bill Freeman.
  https://billf.mit.edu/sites/default/files/documents/cvprPapers.pdf

Thank you!

Before going into questions, I will show/dissect some example good papers.

Questions?