"How to write an award-winning paper"

or "how to be understood by your peers"

Hoel Kervadec

December 31, 2024

Universiteit van Amsterdam
Instituut voor Informatica

Amsterdam UMC

Overview

Why this workshop?

Before writing

Anatomy of a paper

Writing the paper

After submission

Summary and questions

Why this workshop?

Papers are the main way to communicate ideas, to your peers.

Traditional publishing process

- Research & science: you

- Manuscript: you

- Typesetting: professional typesetter

- Printing: professional printer

Modern publishing process

- Research & science: you

- Manuscript: you

- Typesetting: you and LaTeX

- Printing: your responsibility

Who am I?

2017: first reviews (never stopped)

2018: my first MIDL submission, with subsequent journal extension

2020: submit and defend PhD thesis

2021–today: publishing editor at the Melba journal

2023/2024: award committee for MIDL

2024–today: assistant professor
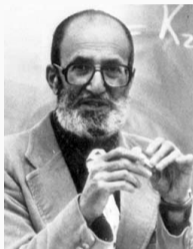
Keep in mind that:

- some papers & books are written over months or even years;
- there is always a deadline, do not burn yourself for one;
- rejection is part of the process.

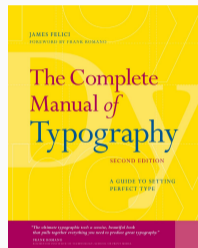How *I* write papers, or actually *how I attempt to write papers*.

Find and refine your own method.

Writing is an exercise of empathy.

"How to write mathematics" — Paul Halmos, 1970

Paul Halmos said:

«The problem is to communicate an idea. To do so, and to do it clearly, you must have something to say, and you must have someone to say it to, you must organize what you want to say, and you must arrange it in the order you want it said in.»

Before writing

Read papers, practice summarizing them

Review papers

How to write an outstanding review

Were you invited to review for MIDL 2025, for another venue or even a journal? Do you want to take your first steps in writing reviews independently, or do you just want to become more efficient in doing so? Join this workshop on 29th January 2025 to learn how to write outstanding reviews. After a tutorial loaded with hints from experienced, senior colleagues, you will discuss real reviews and feedback from previous MIDL editions, including their rating by the program board. You will learn how to open up your own mind to grasp a manuscript from different perspectives, and how a review should be formulated to give value to the authors. We are proud that the workshop will be led by Dr. Lisa Koch from the University of Bern. She will be supported by a great team, stay tuned.
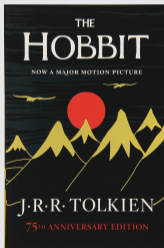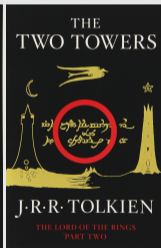
Improve your English

Actively: make a list of interesting formulations.

Passively: Read non-scientific books!

Personal favorites:

Beginner   Intermediate   Advanced

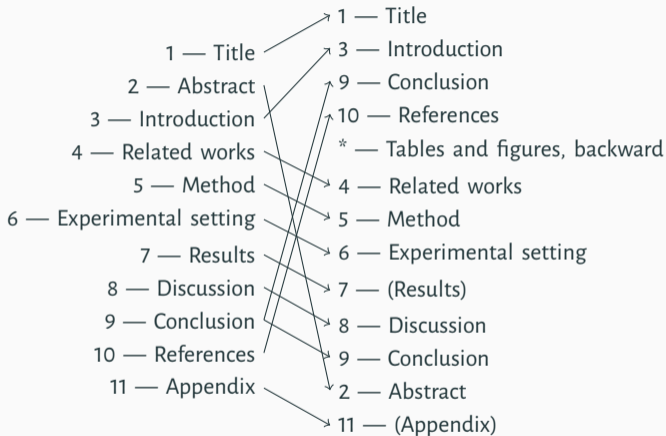Make your life easier, can focus more on the writing.

Include:

- $\LaTeX$;
- text editor;
- automate figures and plots;
- git;
- large screen;
- keyboard and touch-typing.

Check your institute/university policy.

Anatomy of a paper

1 — Title
2 — Abstract
3 — Introduction
4 — Related works
5 — Method
6 — Experimental setting
7 — Results
8 — Discussion
9 — Conclusion
10 — References
11 — Appendix

1 — Title
3 — Introduction
9 — Conclusion
10 — References
* — Tables and figures, backward
4 — Related works
5 — Method
6 — Experimental setting
7 — (Results)
8 — Discussion
9 — Conclusion
2 — Abstract
11 — (Appendix)

Similar structure for benchmark and surveys; though with a different emphasis.

A scientific paper is rarely read from front to back.

From a few minutes to a few hours.

People also come back at it after months or years.

⇒ Information should be easy to find

Short and informative.

How researcher notice your paper in a list.

Synthesis of the paper.

Discuss very quickly:

- topic and motivation;
- overall method (without details);
- main results.

What is the field, why is it of interest.

Which sub-problem do you want to tackle?

How the litterature is not sufficient to solve it.

Overview of your contribution: summary and teaser.

A good figure helps understanding and stoke interest

(a) A visual comparison of the different supervision methods on the ACDC dataset.

Pixel	Label	Shape descriptor	Class		
		(in pixels)	RV	Myo	LV
0	RV	Object volume $\mathfrak{V}$	3100	800	1600
1	Background	Centroid location $\mathfrak{C}$	(125, 80)		(125, 125)
2	LV	Avg. dist. to centroid $\mathfrak{D}$	(20, 15)	(15, 20)	(10, 10)
⋮		Object length $\mathfrak{L}$	750	1000	500
65536	Background				

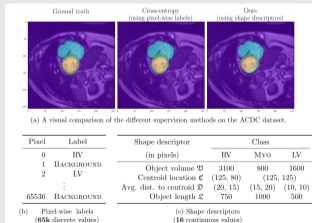(b) Pixel-wise labels (**65k** discrete values) (c) Shape descriptors (**16** continuous values)

Figure 1: RV, Myo and LV stands for "right-ventricle", "myocardium" and "left-ventricle", respectively. The shape descriptors are detailed in Subsection 2.2.

15

Why the existing literature is not sufficient.

Be fair to other researchers works.

Not simply a list: ground-work for your method.

Pseudo-writing:

A did X, and B did Y, but it is not fully applicable to P because of R. Z is the closest work but limited to tasks T.

Most important section: How does your method work?

Supported by:

- mathematical notation (if relevant);
- figures (could be Fig 1).

Some details can be left in the Appendix, or the public code.

Pseudo-writing:

We start from ZZ, which is the simplest and closest model. We build A and B on top of it, because RRR. Our final method is therefore ZZAB.

How are you going to evaluate your method?

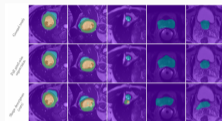Dataset, metrics, main hyperparameters, tools and frameworks.

Methods you compare to (if relevant).

Pseudo-writing:

We use D dataset and evaluate with metrics M1 and M2. Code is implemented in F and is available online. We compare to A because it is relevant.

Both qualitative and quantitative results (if possible).

Method	ACDC				PROMISE12
	RV	MYO	LV	Overall	Prostate
Cross-entropy (pixel-wise)	0.879 (0.066)	0.829 (0.074)	0.919 (0.059)	0.876 (0.076)	0.871 (0.047)
Ours (shape descriptors)	0.825 (0.107)	0.660 (0.114)	0.819 (0.086)	0.768 (0.128)	0.651 (0.098)

Pseudo-writing:

It works, by this much.

What is the consequence of your results?

New findings? New paradigm? Exciting new experiments to perform?

Current limitation? Be honest.

Pseudo-writing:

It works, therefore…

Yet, A, B and C remains.

Summarize your findings:

- which problem did you solve;
- how did you solve it;
- by how much;
- opening: what remains?

Done by LaTeX.

Make sure journals are up to date.

Avoid duplicates.

Start on a `\newpage`: not a mandatory part of the paper.

Put the details not needed to understand the paper.

Writing the paper

First step: Figure out your best writing time slot.

- Start from paper template
- Write sparsely, just fill ideas there and there
- Placeholder for figures and tables
- Ask for first feedback: focus on the big picture and flow, not the details
- Continue filling sections, without priority: as the inspiration comes
- Start filling results, and see if there is a mismatch with the current writing
- Start refining from beginning to end. Iterate
- Near the deadline, compress the paper: remove redundant parts, make the sentences more to the point, tweak the figures placement

Refine a method that works for you. Make a writing plan, or not.

Revisiting bounding boxes: weakly supervised image segmentation with inequality constraints and tightness prior

Anonymous Dupont

Abstract

This is a great paper and it has a concise abstract.
Keywords: CNN,image segmentation,weak supervision,bounding-boxes

1. Introduction

Bounding boxes: uncertainty to deal with
 Can reuse annotations made for object detection

2. Related works

Deep cut and Papandreou et al
 Grab cut ?
 That NIPS 2019
 Orig 2009 from victor

3. Method

Put here a Figure with the labels used

3.1. On how to deal with the certainty outside the box

3.2. On how to deal with the uncertainty inside the box

3.3. Additional regularizarion: contraining the size

3.4. Constrained optimization with log-barrier

3.5. Final model

4. Experiments

4.1. Datasets

4.1.1. PROSTATE SEGMENTATION ON MR-T2

The third dataset was made available at the MICCAI 2012 prostate MR segmentation challenge[1]. It contains the transversal T2-weighted MR images of 50 patients acquired at different centers with multiple MRI vendors and different scanning protocols. It is comprised of various diseases, i.e., benign and prostate cancers. The images resolution ranges from $15 \times 256 \times 256$ to $54 \times 512 \times 512$ voxels with a spacing ranging from $2 \times 0.27 \times 0.27$ to $4 \times 0.75 \times 0.75mm^3$. We employed 40 patients for training and 10 for validation.

1. https://promise12.grand-challenge.org

4.1.2. ATLAS

Brain lesion, 100 patients

4.2. Implementation details

Fully residual Unet for PROMISE with BS 4, ENet for ATLAS with BS 32
 Data aug for promise
 Batch the box prior constraints by 5 ; this value didn't affect much the results
 log-barrier params: defaults value from logbarrier paper
 Balance of the weights can be found in the recipe in the code (prostate.make, atlas.make).

5. Results

Method	3D DSC
Deep cut	
Box prior + box size	---
w/ negative cross-entropy	77.45
w/ empty background size	81.66
Full supervision (Cross-entropy)	90.09

Table 1: Table of the results for PROMISE12

Method	3D DSC	2D DSC	HD
Deep cut	-	-	-
Box prior + box size			
w/ negative cross-entropy	-	-	-
w/ empty background size	-	-	
Full supervision (Cross-entropy)	-	-	-

Table 2: Table of the results for ATLAS

5.1. Efficiency

On a machine with a Titan RTX:

Method	it/s (PROMISE)	it/s (ATLAS)
Base (cross-entropy)	9	5
w/ box prior (naive implementation)	7	
w/ box prior (fast implementation)	8?	
Deep cut	??	??

Table 3: Comparison in training speed between the different methods

Page limit

Do not be concerned with it initially.

It is easy to "compress" 1 or 2 extra pages near the deadline.

Write in plain English, use your own vocabulary.

Do not use short ways ("Don't", "Isn't", …).

Remember: typographic conventions might be different from your native language.

If possible, ask for proof-reading from a native speaker.

Ideal writing is spread over weeks, even months.

Some tips to accelerate it:

- Sleep!
- Exercise
- Any social activity <u>bound</u> in time
- Print the paper, draft with a pen
- Use a different pdf viewer/change the zoom settings
- Read it out loud

« Change is the master key. [...] It is not enough merely to switch off the lights which play upon the main and ordinary field of interest; a new field of interest must be illuminated. It is no use saying to the tired 'mental muscles'[...] 'I will give you a good rest' [...]. The mind keeps busy just the same. »

— Winston Churchill (Painting as a pastime, 1948)

Paul Halmos said:

«The writer must anticipate and avoid the reader's difficulties. As he writes, he must keep trying to imagine what in the words being written may tend to mislead the reader, and what will set him right.»

Do not start with the final model, tell a story.

Be honest: how did you come up with your method?

Why did you do *this*, and not *that*?

Paul Halmos said:

«Half of the art of good writing is the art of omission.»

Do not drown the reader with details. Reintroduce what is needed, without copying the full bibliography

Rule of thumb: how long did it take for you to notice it was required?

A picture is worth a thousand words.

Do not try to sound smart.

The *reader* should feel smart.

Gently nudge him toward the next idea: make him "guess" it.

Nicolas Boileau-Despréaux said:

« Whatever is well conceived is clearly said,
  And the words to say it flow with ease. »

If you struggle to describe your idea, perhaps should go back on the drawing board.

Is your method as simple as it could be? Are you clear on what you are trying to achieve?

Do you have multiple goals and ideas that could be spread in several papers?

A picture is worth a thousand words...

...but at first take a thousand more time to make.

Rule of thumb: half a day per good figure.

Develop over your PhD: https://hkervadec.github.io/pyplot_zoo/

Polish a bit more at each submission.

Figures (ii)

Automate your figures as much as possible.

Many frameworks: Matplotlib, Plotly, Tikz (in LaTeX), Manim, ...

Can finish/polish with Photoshop, Inkscape, ...

Time invested in tools always pays back.

Export figures to PDF: make text selectable.

Paul Halmos said:

«[One] other thing I would recommend that you do first is to invest an hour or two of thought in the alphabet ; you will find it saves many headaches later.»

Styles help:

`$abcABC123$`	$abcABC123$
`$\mathbf{abcABC123}$`	$\mathbf{abcABC123}$
`$\mathsf{abcABC123}$`	$\mathsf{abcABC123}$
`$\mathfrak{abcABC123}$`	$\mathfrak{abcABC123}$
`$\mathcal{ABC}$`	$\mathcal{ABC}$
`$\mathbb{ABC}$`	$\mathbb{ABC}$

Font styles can carry semantics.

Make it public if you can: https://www.midl.io/reproducibility

Can reference it in the manuscript.

https://zenodo.org/ good option for archival and citation.

Prose is not mathematics which is not code.

They all complement each others.

Ask regularly for feedback, from outsider.

What kind of feedback you need:

- overall sections structure;
- flow of ideas;
- formulation of the method;
- English and grammar;
- a specific Section?

Learn to accept criticism.

Build your network of proof-readers.

What about conflicting feedback?

Take the person's background into account

Different fields have different styles and idiosyncrasies.

Ultimately, you are the first author:

it is *your* paper(, your PhD).

Mention the good.

Be constructive: suggest improvements.

Mention if something is *difficult* to understand.

Start compressing the paper:

- (shorter) sentences more to the point;
- discard what is (in hindsight) less relevant;
- tune figure placements;
- place figures close to their referring text.

Writing helpers

Table of contents (with deep sectioning).

`\usepackage[pass, showframe]{geometry}`

Indent your code.

Commit often, keep your `.tex` clean.

Do not modify a sentence, rewrite it from scratch.

After submission

Clean-up, archival

Clean-up your code, remove comments, fix indentation.

Archive the:

- latex code;
- results and training artifacts.

Redo your figures with the better/cleaner framework?

Get ready...

Rebuttal or revision are coming:

- new metrics to compute;
- new experiments to run;
- merging results from different runs.

Summary and questions

Carl Friedrich Gauss said:

«I write slowly. This is chiefly because I am never satisfied until I have said as much as possible in a few words, and writing briefly takes far more time than writing at length.»

But what is an award winning paper?

- a good idea;
- well executed;
- *and clearly communicated.*

Award winning papers stand-out on first glance, *no need to dig deep* to realize it is good.

Summary

Main ingredient is time

Find a way to make writing fun, *for you*

Start early...

... and iterate

and again.