UNIVERSITY OF CALIFORNIA, MERCED

How to Get Your CVPR Paper Rejected?

Ming-Hsuan Yang

UCMERCED

www.ucmerced.edu

Outline

- Conferences
- Journals
- Writing
- Presentation
- Lessons

Conferences

- CVPR – Computer Vision and Pattern Recognition, since 1983
  - Annual, held in US
- ICCV – International Conference on Computer Vision, since 1987
  - Every other year, alternate in 3 continents
- ECCV – European Conference on Computer Vision, since 1990
  - Every other year, held in Europe

Conferences

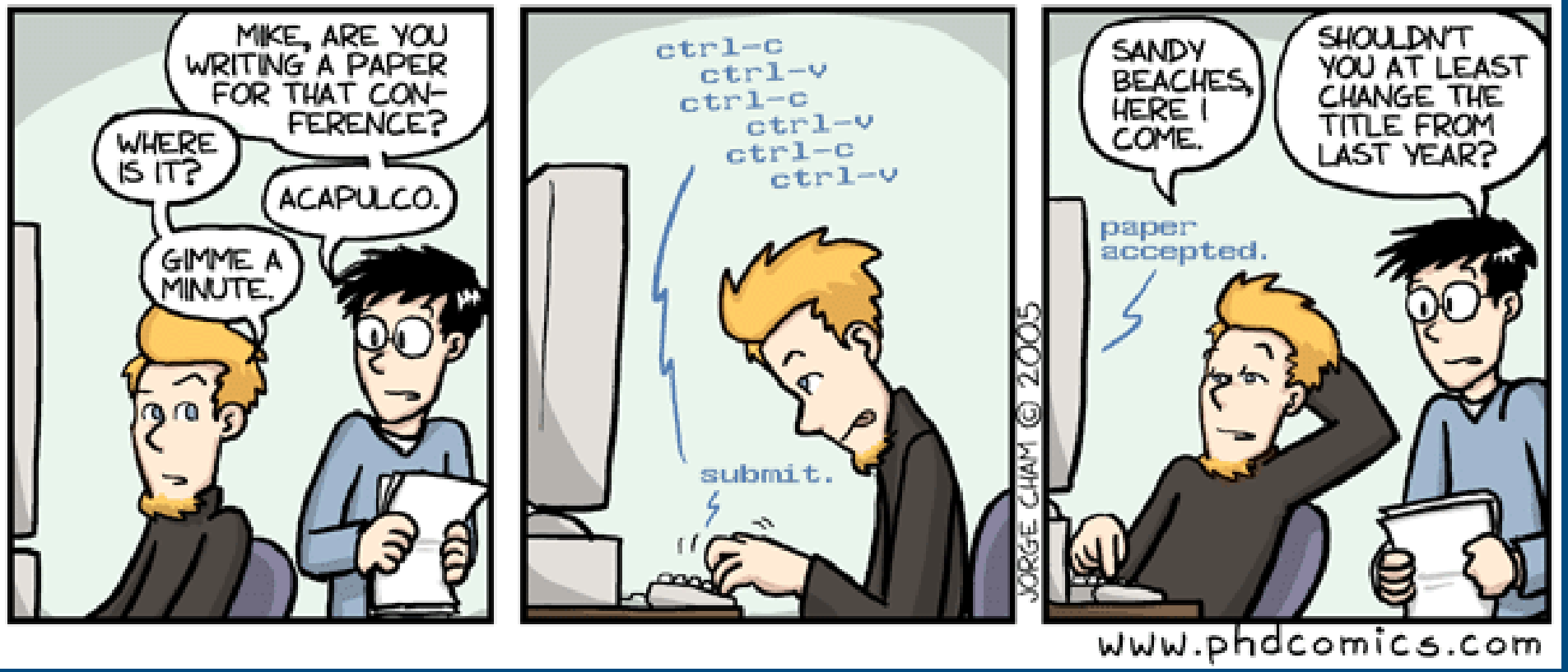
- ACCV – Asian Conference on Computer Vision
- BMVC – British Machine Vision Conference
- ICPR – International Conference on Pattern Recognition
- SIGGRAPH
- NIPS – Neural Information Processing Systems

Conferences

- MICCAI – Medical Image Computing and Computer-Assisted Intervention
- FG – IEEE Conference on Automatic Face and Gesture Recognition
- ICCP – IEEE International Conference on Computational Photography
- ICML – International Conference on Machine Learning
- IJCAI, AAAI, MVA, ICDR, ICVS, DAGM, CAIP, ICRA, ICASSP, ICIP, SPIE, DCC, WACV, 3DPVT, ACM Multimedia, ICME, …

Conference Location

Conference Location



- Me and confernece I want to attend (location vs. reputation)

Conference Organization

- General chairs: administration
- Program chairs: handling papers
- Area chairs:
  - Assign reviewers
  - Read reviews and rebuttals
  - Consolidation reports
  - Recommendation
- Reviewers
- Authors

Review Process

- Submission
- CVPR/ECCV/ICCV
  - Double blind review
  - Program chairs: assign papers to area chairs
  - Area chairs: assign papers to reviewers
- Rebuttal
- Results

Area Chair Meetings

- Each paper is reviewed by 2/3 area chairs
- Area chair make recommendations
- Program chairs make final decisions
- Virtual meetings
- Onsite meetings
  - Several panels
  - Buddy/triplet

Triage

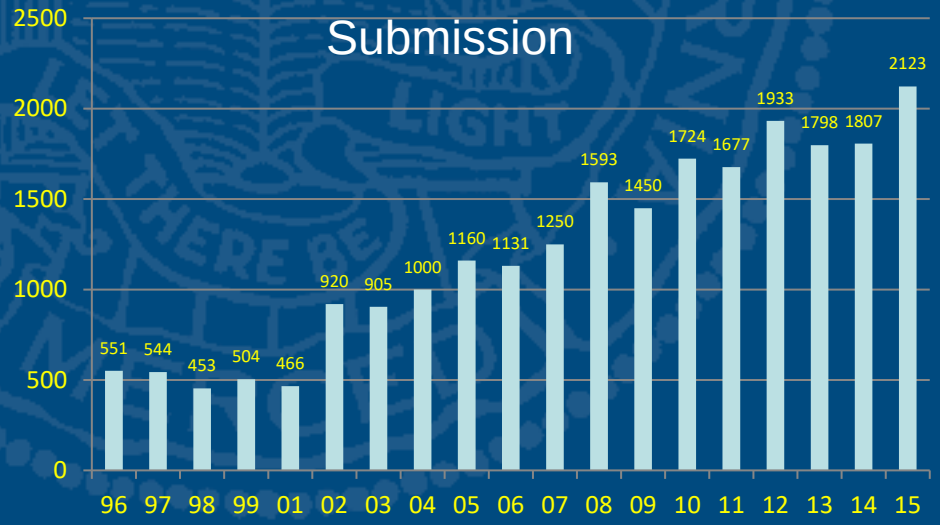
- Area chairs know the reviewers
- Reviews are weighted
- Based on reviews and rebuttal
  - Accept: (decide oral later)
  - Reject: don't waste time
  - Go either way: lots of papers
- Usually agree with reviewers but anything can happen as long as there are good justifications

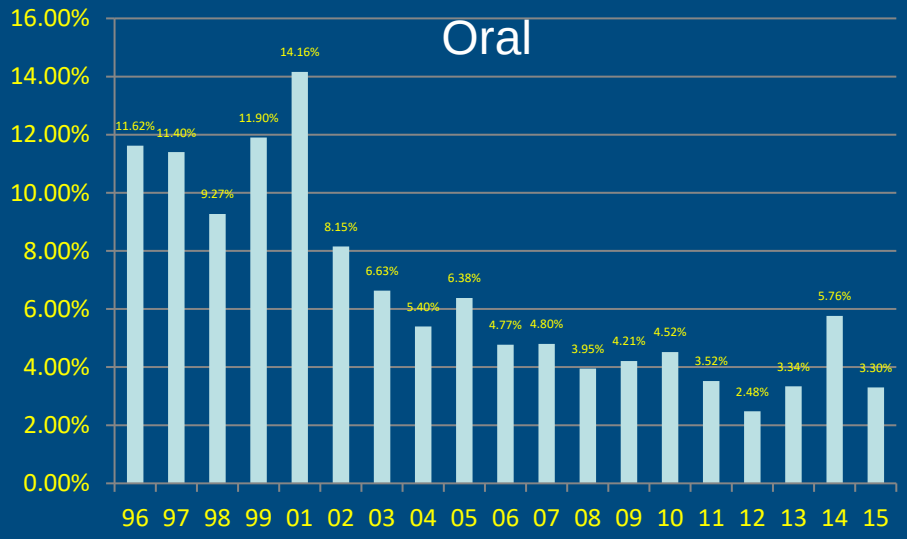
Conference Acceptance Rate

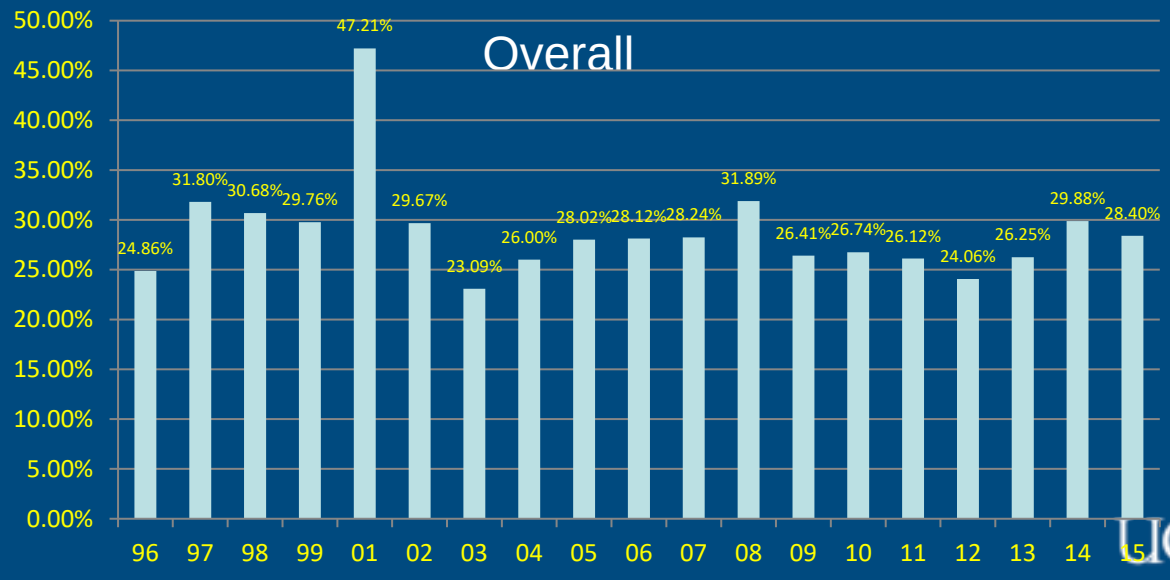
- ICCV/CVPR/ECCV: ~ 25%
- ACCV (2009): ~ 30%
- NIPS: ~ 25%
- BMVC: ~ 30%
- ICIP: ~ 45%
- ICPR: ~ 55%

- Disclaimer
  - low acceptance rate = high quality?

CVPR

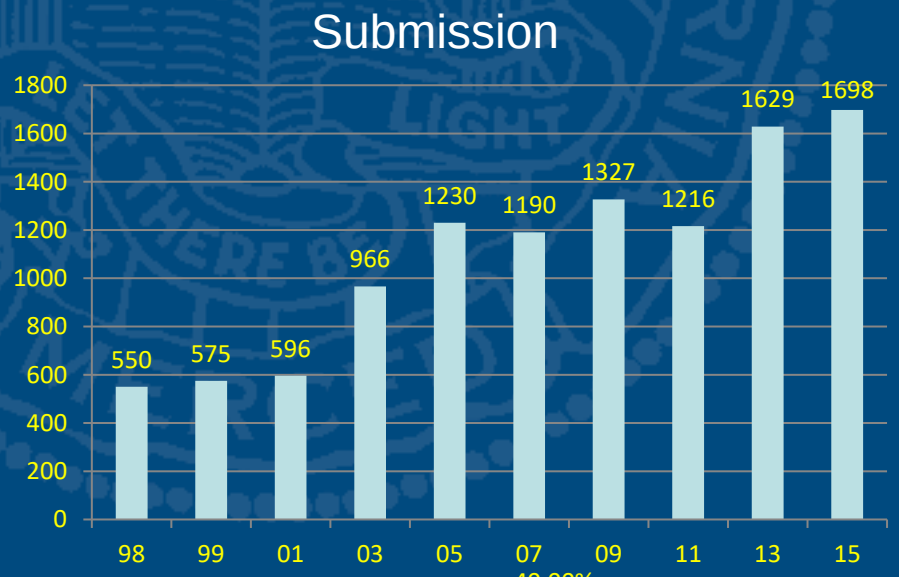


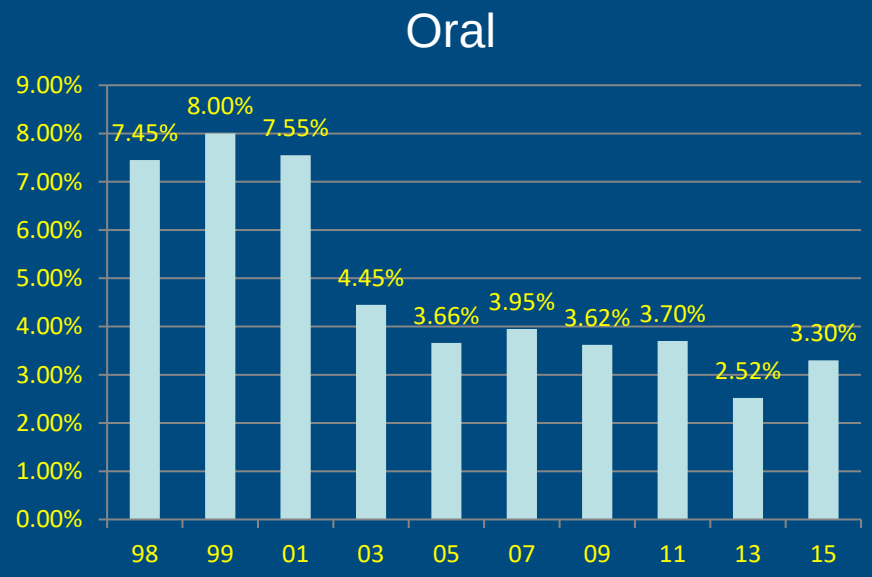


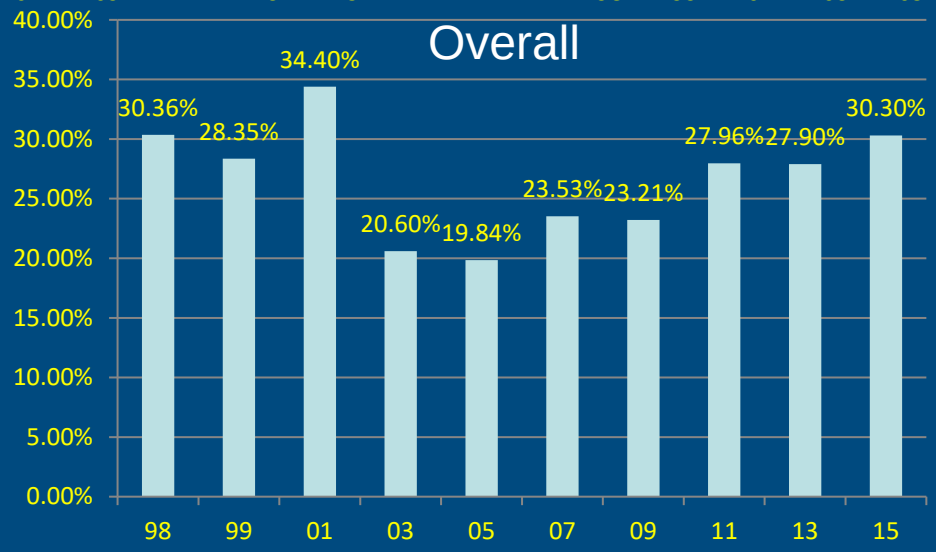


UCMERCED

ICCV

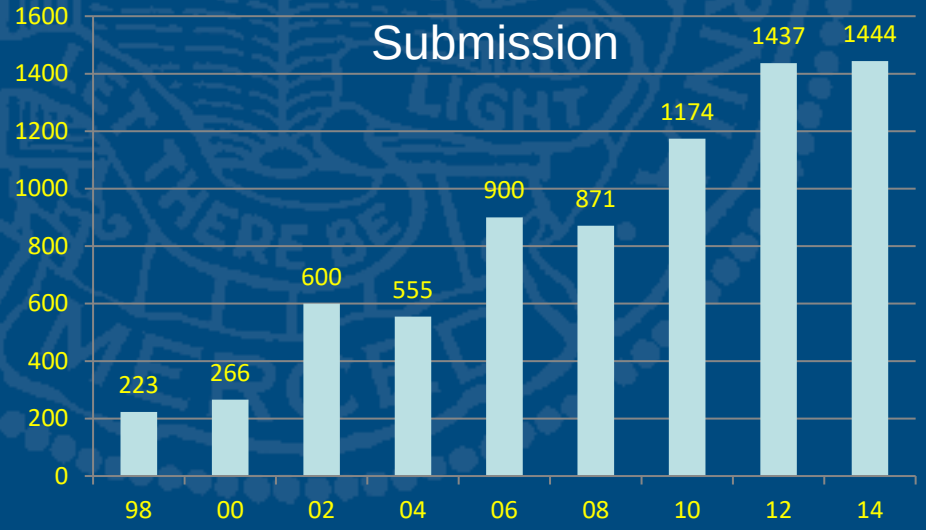


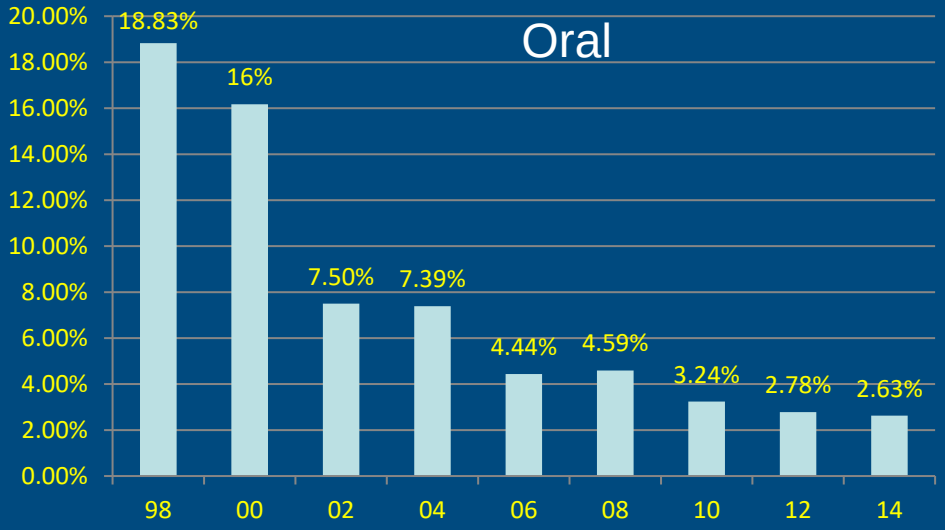


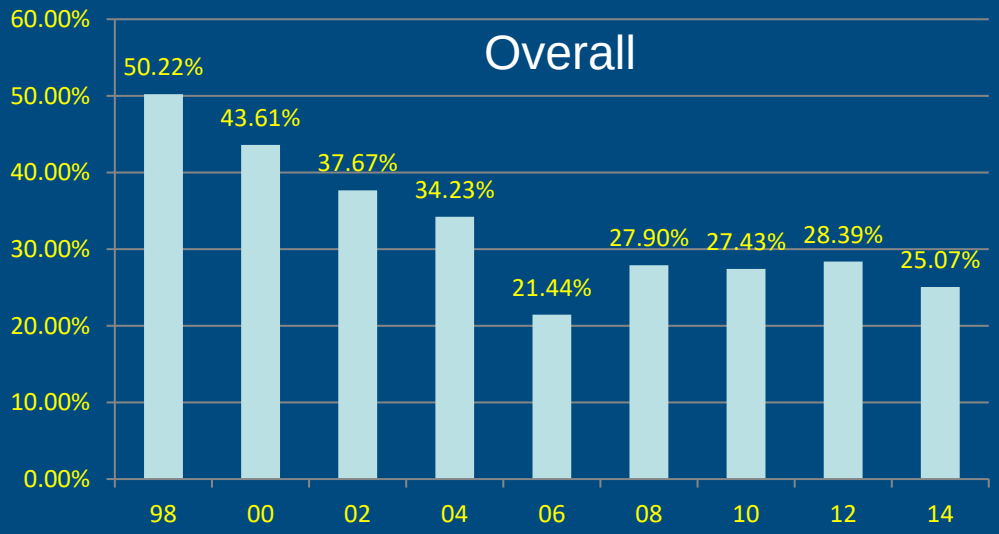


UCMERCED

ECCV







UCMERCED

Top 100 Publications - English

- For what it is worth (h5 index by Google Scholar)

1. Nature
2. The New England Journal of Medicine
3. Science

…

55. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

…

Top Publications - E&CS

1. Nano Letters

…

8. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

…

16. IEEE Transactions on Pattern Analysis and Machine Intelligence

…

Reactions

- Top journal papers
- Workshops vs conferences
- Waiting for the review or final results
- Acceptance
- Reject
- Mixed feeling
- Finding an error
- Resubmit?
- This time, it will go through
- Paper finally accepted
- Registration
- Oral presentation
- Poster presentation

Database Community

- Jeffrey Naughton's ICDE 2010 keynote
- What's wrong with the reviewing process?
- How to fix that?

Journals

- PAMI – IEEE Transactions on Pattern Analysis and Machine Intelligence, since 1979 (impact factor: 5.96, #1 in all engineering and AI, top-ranked IEEE and CS journal)
- IJCV – International Journal on Computer Vision, since 1988 (impact factor: 5.36, #2 in all engineering and AI)
- CVIU – Computer Vision and Image Understanding, since 1972 (impact factor: 2.20)

Journals

- IVC – Image and Vision Computing
- TIP – IEEE Transactions on Image Processing
- TMI- IEEE Transactions on Medical Imaging
- MVA – Machine Vision and Applications
- PR – Pattern Recognition
- TMM – IEEE Transactions on Multimedia
- …

PAMI Reviewing Process

- Associate editor-in-chief (AEIC) assigns papers to associate editors (AE)
- AE assigns reviewers
- First-round review: 2-4 months
  - Accept as is
  - Accept with minor revision
  - Major revision
  - Resubmit as new
  - Reject

PAMI Reviewing Process

- Second-round review: 2-4 months
  - Accept as is
  - Accept with minor revision
  - Major revision (rare cases)
  - Reject
- EIC makes final decision
- Overall turn-around time: 6 to 12 months
- Rule of thumb: 30% additional work beyond a CVPR/ICCV/ECCV paper

IJCV/CVIU Reviewing Process

- Similar formats
- Slightly longer turn-around time

Journal Acceptance Rate

- PAMI
  - 2013: 151/959: 15.7%
  - 2014: 160/1018: 15.7%
- IJCV: ~ 20% (my guess, no stats)
- CVIU: ~ 25% (my guess, no stats)

From Conferences to Journals

- How much additional work?
  - 30% additional more work for PAMI?
  - As long as the journal version is significantly different from the conference one
- Novelty of each work
  - Some reviewers still argue against this
  - Editors usually accept paper with the same ideas

How to Get Your CVPR Paper Rejected?

- Jim Kajia (SIGGRAPH 93 papers chair): How to get your SIGGRAPH paper rejected?
- Bill Freeman: How to write a good CVPR submission
- Do not
  - Pay attention to review process
  - Put yourself as a reviewer to exam your work from that perspective
  - Put the work in right context
  - Carry out sufficient amount of experiments
  - Compare with state-of-the-art algorithms
  - Pay attention to writing

Review Form

- Summary
- Overall Rating
  - Definite accept, weakly accept, borderline, weakly reject, definite reject
- Novelty
  - Very original, original, minor originality, has been done before
- Importance/relevance
  - Of broad interest, interesting to a subarea, interesting only to a small number of attendees, out of CVPR scope

Review Form

- Clarity of presentation
  - Reads very well, is clear enough, difficult to read, unreadable
- Technical correctness
  - Definite correct, probably correct but did not check completely, contains rectifiable errors, has major problems
- Experimental validation
  - Excellent validation or N/A (a theoretical paper), limited but convincing, lacking in some aspects, insufficient validation
- Additional comments
- Reviewer's name

Learn from Reviewing Process

- Learn how others/you can pick apart a paper
- Learn from other's mistakes
- Get to see other reviewers evaluate the same paper
- See how authors rebut comments
- Learn how to write good papers
- Learn what it takes to get a paper published

Put Yourself as Reviewer

- Reviewer's perspective
- How a paper gets rejected?
- What are the contributions?
- Does it advance the science in the field?
- Why you should accept this paper?
- Is this paper a case study?
- Is this paper interesting?
- Who is the audience?

Novelty

- What is new in this work?
  - Higher accuracy, significant speed-up, scale-up, ease to implement, generalization, wide application domain, connection among seemingly unrelated topics, ...
- What are the contributions (over prior art)?
- Make a compelling case with strong supporting evidence

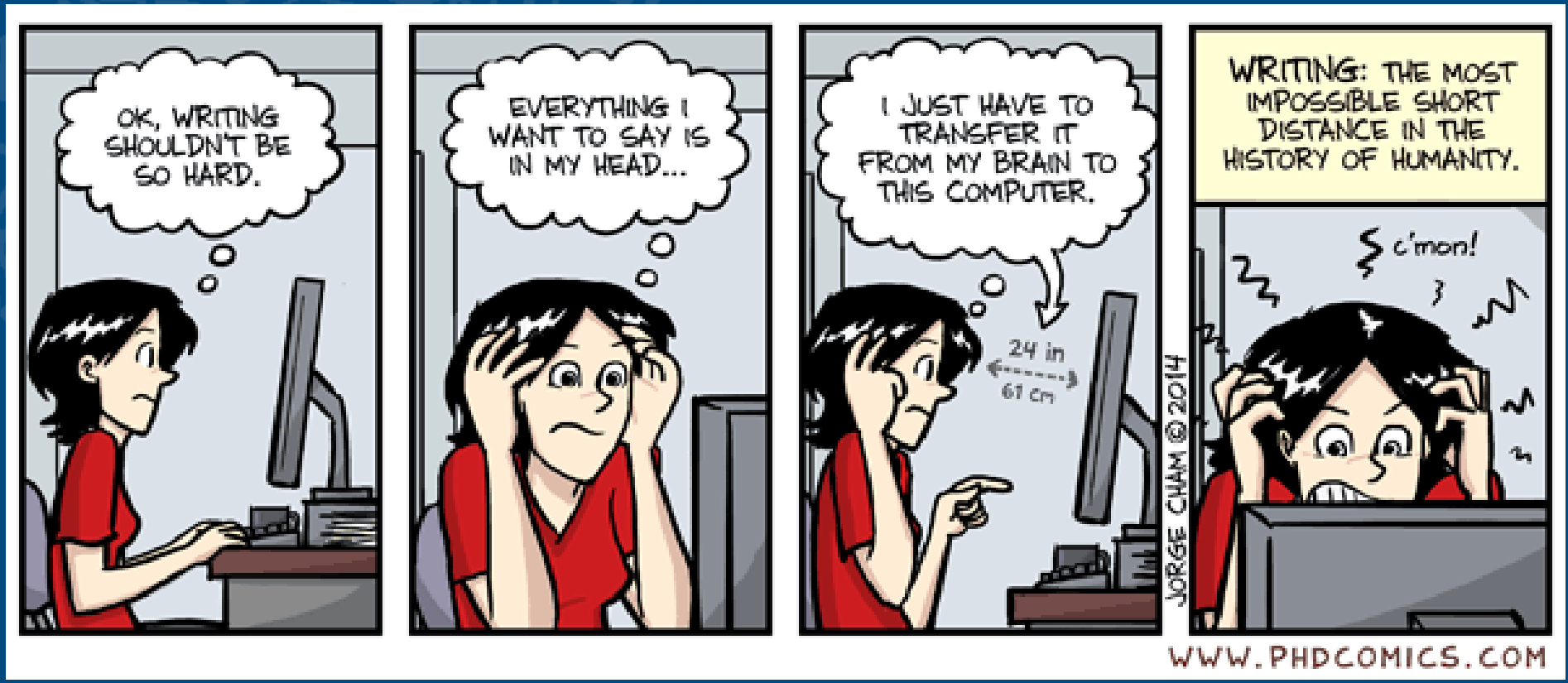
Experimental Validation

- Common data set
- Baseline experiment
- Killer data set
- Large scale experiment
- Evaluation metric
- Realize things after submission
- Friendly fire

Compare With State of the Art

- Do your homework
- Need to know what is out there (and vice versa)
- Need to show why one's method outperforms others, and in what way?
  - speed?
  - accuracy?
  - sensitive to parameters?
  - assumption
  - easy to implement?
  - general application?

Writing


OK, WRITING SHOULDN'T BE SO HARD.
EVERYTHING I WANT TO SAY IS IN MY HEAD...
I JUST HAVE TO TRANSFER IT FROM MY BRAIN TO THIS COMPUTER.
24 in
61 cm
WRITING: THE MOST IMPOSSIBLE SHORT DISTANCE IN THE HISTORY OF HUMANITY.
c'mon!
JORGE CHAM © 2014
WWW.PHDCOMICS.COM

Writing

OK, I NEED TO TRANSFER MY THOUGHTS FROM MY HEAD TO THIS SCREEN...
C'MON, WRITE!
WRITE! WRITE!
JORGE CHAM © 2014
THEY'RE TOUCHING! TRANSMIT! TRANSMIT!
thud! thud!
WWW.PHDCOMICS.COM

Writing

- Reviewing a poorly written paper
- Clear presentation
- Terse
- Careful about wording
- Make claims with strong evidence

Writing

- Matt Welsh's blog on scientific writing
- Sharpen your mental focus
- Force you to obsess over every meticulous detail – word choice, word count, overall tone, readability of graphs (and others such as font size, layout and spacing, and page limit)

Writing

- Crystalizing the ideas through the process of putting things together
- Hone the paper to a razor-sharp, articulate, polished work

Writing

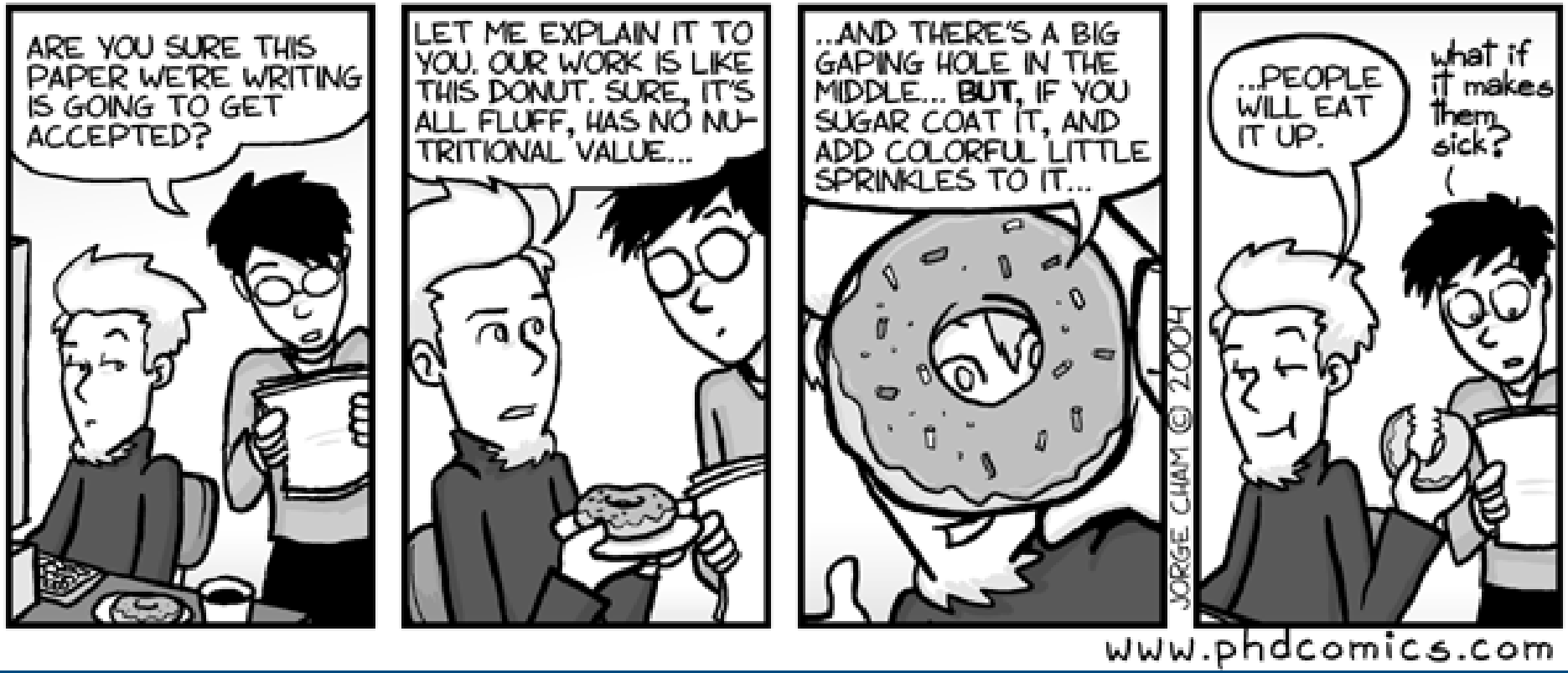
- Write the paper *as early as possible*, sometimes *before* even starting the research work
- Will discover the important things that you have not thought about
- The process of writing results in a flood of ideas

Writing

- Even if a paper is not accepted, the process is energizing and often lead to new ideas for the next research problems
- Submitting the paper is often the start of a new line of work
- Riding on that clarity of thought would emerge post-deadline (and a much-needed break)

Tell A Good Story

- Good ideas and convincing results
- But not too much (vs grant proposal)



Presentation

- Good artists copy, great artists steal
- Not just sugar coating
- Not just a good spin
- Tell a convincing story with solid evidence
- Present your ideas with style
- Q&A
- Real stories

Interesting Title

- Cool titles attract people
- Grab people's attention
- Buzz word?
- But don't be provocative

Math Equations

- Minimal number of equations
  - No more, no less
  - Too many details simply make a paper inaccessible
- Too few equations
- Many good papers have no or few equations
  - CVPR 13 best paper
  - CVPR 05 HOG paper

Figures

- Be clear
- Sufficient number of figures

Theoretical or Applied?

- Computer vision is more applied, at least nowadays
- Theory vs real world
- More high impact papers are about how to get things done right

Common Mistakes

- Typos
- Unsupported claims
- Unnecessary adjectives (superior!)
- “a”, “the”
- Inanimate objects with verbs
- Inconsistent usage of words
- Laundry list of related work (or worse copy sentences from abstracts)
- Bad references
- Laundry list of related work
- Repeated boring statements

Get Results First than Writing?

- Conventional mode
  - Idea-> Do research -> Write paper
- "How to write a great research paper" by Simon Peyton Jones
  - Idea -> Write paper -> Do research
    - Forces us to be clear, focused
    - Crystallizes what we don't understand
    - Opens the way to dialogue with others: reality check, critique, and collaboration
- My take
  - Idea -> Write paper -> Do research -> Revise paper -> Do research -> Revise paper -> …

Supplementary Material

- Important
- Add more results and large figures
- Add technical details as necessary (don't miss important details)
- Derivation details, e.g., proof of a theorem

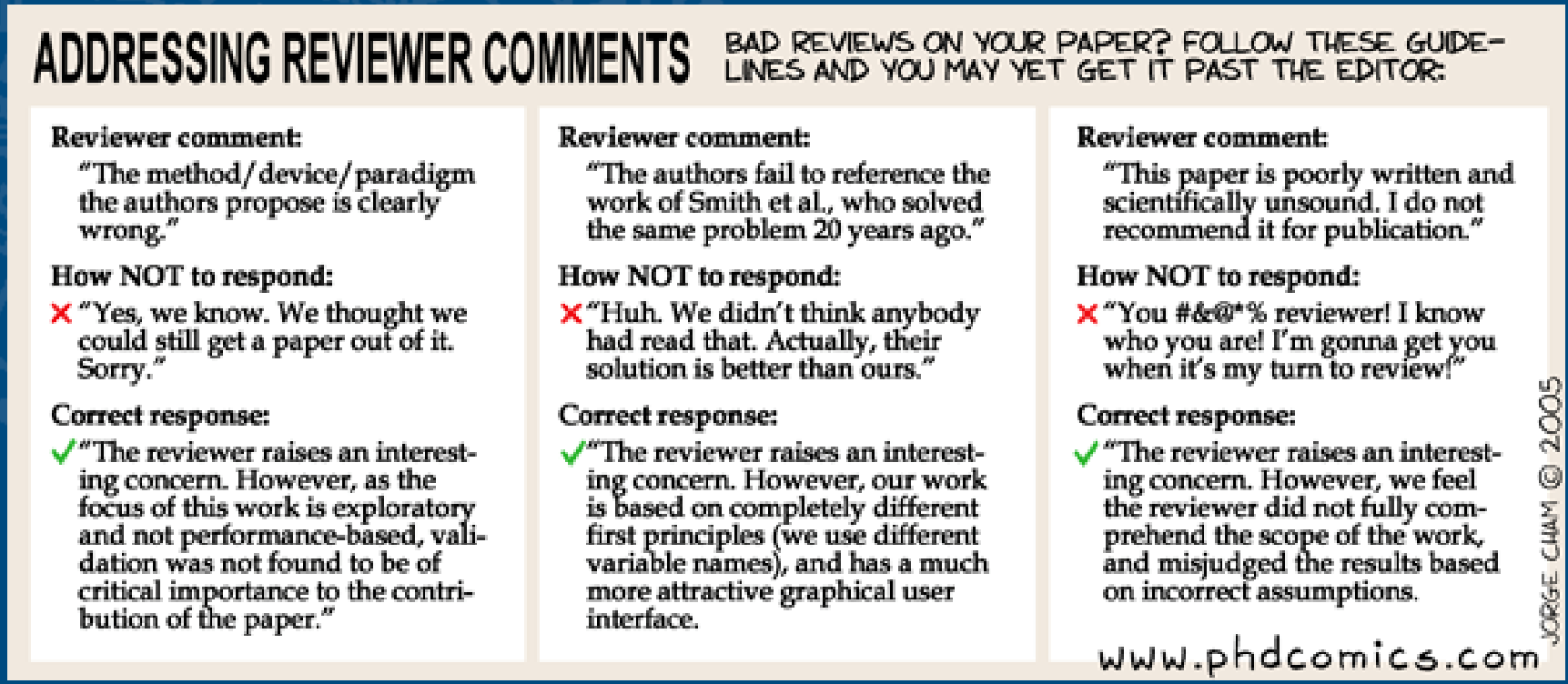
Most Important Factors

- Novelty
- Significant contributions (vs. salami publishing)
- Make sure your paper is non-rejectable (above the bar with some error margin)

Reviews

- Me: Here is a faster horse
- R1: You should have used my donkey
- R2: This is not a horse, it’s a mule
- R3: I want a unicorn!

Rebuttal or Response



ADDRESSING REVIEWER COMMENTS

BAD REVIEWS ON YOUR PAPER? FOLLOW THESE GUIDELINES AND YOU MAY YET GET IT PAST THE EDITOR:

Reviewer comment:
"The method/device/paradigm the authors propose is clearly wrong."

How NOT to respond:
✗ "Yes, we know. We thought we could still get a paper out of it. Sorry."

Correct response:
✓ "The reviewer raises an interesting concern. However, as the focus of this work is exploratory and not performance-based, validation was not found to be of critical importance to the contribution of the paper."

Reviewer comment:
"The authors fail to reference the work of Smith et al., who solved the same problem 20 years ago."

How NOT to respond:
✗ "Huh. We didn't think anybody had read that. Actually, their solution is better than ours."

Correct response:
✓ "The reviewer raises an interesting concern. However, our work is based on completely different first principles (we use different variable names), and has a much more attractive graphical user interface.

Reviewer comment:
"This paper is poorly written and scientifically unsound. I do not recommend it for publication."

How NOT to respond:
✗ "You #&@*% reviewer! I know who you are! I'm gonna get you when it's my turn to review!"

Correct response:
✓ "The reviewer raises an interesting concern. However, we feel the reviewer did not fully comprehend the scope of the work, and misjudged the results based on incorrect assumptions.

JORGE CHAM © 2005

www.phdcomics.com

Good surprise	Bad surprise
• One CVPR paper: BR, BR, DR • Two ECCV paper: PR, PR, BR • One CVPR 15 paper: BR, BR, WR -> poster, poster, poster • One CVPR 15 paper: DR, WA, BR -> Poster, Poster, WR	Two ECCV papers: PA, PA, BR One CVPR 15 paper: WA, BR, BR -> Poster, Poster, WR One CVPR 16 paper: WR, WR, BR

UCMERCED

Never Know What will Happen

Masked Meta-Reviewer ID: Meta_Reviewer_1
Meta-Reviews:
Question
Consolidation Report
All reviewers agree that this paper has moderate novelty of using partial and spatial information for sparse representation. However, they also concern about
- unclear presentation on technical details (eg. definitions, inference algorithm, pooling methods, template updating schemes, experimental settings etc.),
- not extensive experimental comparison (needs tests on more challenging videos),
- missing justification of the assumption (complementary nature of two kinds of pooling features) and the efficacy of each term.

The authors rebuttal addresses most issues, but is not sufficient to ease the main concerns of R1 and R2. So, the AC recommends the paper to be rejected as it is.

Decision
Definitely Accept

Challenging Issues

- Large scale
  - CVPR 2011 best paper: pose estimation
  - CVPR 2013 best paper: object detection
- Unconstrained
- Real-time
  - CVPR 2001: face detector
  - CVPR 2006: scalable object recognition
- Robustness
- Recover from failure

Interesting Stats

- Best papers and top cited papers in computer science
- Best papers = high impact?
- Oral papers are more influential?
- CVPR Longuet-Hggins prize
- ICCV Helmholtz award

Data Set Selection

- NIPS 02 by Doudou LaLoudouana and Mambobo Bonouliqui Tarare, Lupano Tecallonou Center, Selacie, Guana
- The secret to publish a paper in machine learning conferences?
- Read the references therein carefully!

Data Set Selection

Data Set Selection

Doudou LaLoudouana* and Mambobo Bonouliqui Tarare
Lujaza Tecalkuau Center
Selacie, GUANA
doudoula3@hotmail.com, fuzzybear@yahoo.com

Abstract

We introduce the community to a new construction principle whose practical implications are very broad. Central to this research is the idea to improve the presentation of algorithms in the literature and to make them more appealing. We define a new notion of capacity for data sets and derive a methodology for selecting from them. The experiments show that even for not so good algorithms, you can show that they are significantly better than all the others. We give some experimental results, which are very promising.

1 Introduction

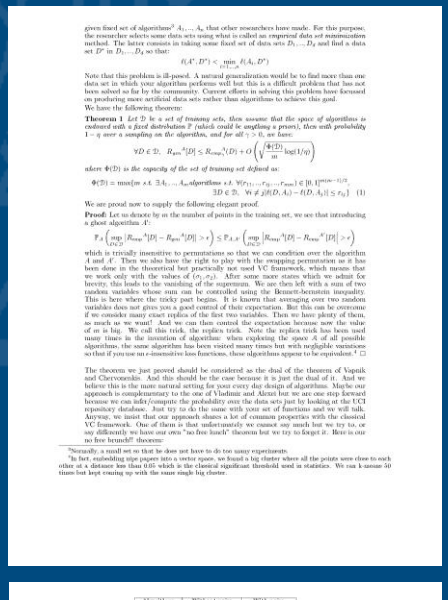
Learning is a marvellous subject. A year spent in artificial intelligence is enough to make one believe in God. Unfortunately, so far it has been handled only from a particular one-sided point of view. The VC-theory known only by some people does not offer what we would be in right to ask from such a theory: we want good bound for our algorithms. We offer with this article a brand new approach that allows you to present your algorithm in a much more principled and rigorous way that has been done before. Many researchers, especially in publications at NIPS, have tried to show when their algorithms are better (in some sense) than some other given set of algorithms. To do this they have employed techniques of data set selection. It is strange then that learning theorists, as they call themselves, over the last 50 years have concentrated on the model selection problem and not the data selection problem which is what people actually do. The two problems can in some sense be seen as the dual of each other, but it is not because you solve one that you can solve the other one. And vice-versa. In this article we lay down the foundations and introduce to the community of machine learning peers and other engineers a new induction principle: structural dataset minimization. Essentially we begin to formalize the sometimes ad hoc engineering approach of the selection procedure that everyone already practiced. In doing so we find concrete bounds for when the data selected really is better than other datasets and implement less ad hoc algorithms for finding such datasets. We show our approach outperforms the classical approach.

The structure of the paper contrarily to its content follows a classical trend: section 1 presents some nice bounds you can use in lots of situations, section 2 shows how to use these bounds by designing new algorithms. Section 3 describe some experiments which, of course, are good.[1] Section 4 concludes the article with smart thoughts and future work we will do.[2]

2 Bounds

Let us introduce some notations. Assume a researcher has invented an algorithm A^* and he wishes to show that his pride and joy is superior with respect to some loss function ℓ to a

[1] Contrarily to other people, we have put all of our experiments in this paper, even the bad ones, but well, we did not get any bad ones.
[2] If (and only if) the paper is accepted.



given fixed set of algorithms[3] $A_1, ..., A_n$ that other researchers have made. For this purpose, the researcher selects some data sets using what is called an *empirical data set minimization* method. The latter consists in taking some fixed set of data sets $D_1, ..., D_S$ and find a data set D^* in $D_1, ..., D_S$ so that:

$$\ell(A^*, D^*) < \min_{D \in D_1,..,D_S} \ell(A_i, D^*)$$

Note that this problem is ill-posed. A natural generalization would be to find more than one data set in which your algorithm performs well but this is a difficult problem that has not been solved so far by the community. Current effort in solving this problem have focused on producing more artificial data sets rather than algorithms to achieve this goal.
We have the following theorem:

Theorem 1 *Let $\mathcal{D}$ be a set of training sets, then assume that the space of algorithms is endowed with a fixed distribution $\mathbb{P}$ (which could be anything a priori), then with probability $1-q$ over a sampling on the algorithm, and for all $\gamma > 0$, we have:*

$$\forall D \in \mathcal{D}, \quad R_{gen}{}^{A}[D] \leq R_{emp}{}^{A}[D] + O\left(\sqrt{\frac{\Phi(\mathcal{D})}{m}\log(1/q)}\right)$$

where $\Phi(D)$ is the capacity of the set of training set defined as:

$$\Phi(\mathcal{D}) = \max\{m \text{ s.t. } \exists A_1, ..., A_m \text{ algorithms } \epsilon A \ \forall (\tau_{11}, ..., \tau_{1l}, ..., \tau_{ml}) \in [0,1]^{ml(m-1)/2}, \\ \exists D \in \mathcal{D}, \ \forall i \neq j | \ell(D, A_i) - \ell(D, A_j)| \leq \tau_{ij}\} \quad (1)$$

We are proud now to supply the following elegant proof.

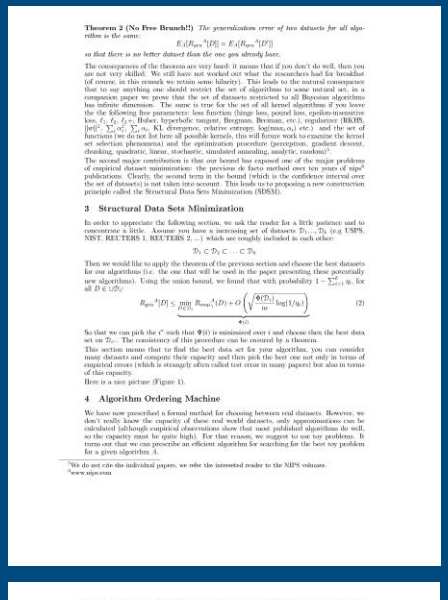
Proof: Let us denote by m the number of points in the training set, we see that introducing a *ghost* algorithm A':

$$\mathbb{P}_A\left(\sup_{\ell \epsilon \mathcal{L}(\mathcal{D})} |R_{emp}{}^{A}[D] - R_{gen}{}^{A}[D]| > \epsilon\right) \leq \mathbb{P}_{A,A'}\left(\sup_{\ell \epsilon \mathcal{L}(\mathcal{D})} \left|R_{emp}{}^{A}[D] - R_{emp}{}^{A'}[D]\right| > \epsilon\right)$$

which is trivially insensitive to permutations so that we can condition over the algorithm A and A'. Then we also have the right to play with the swapping permutation as it has been done in the theoretical but practically not used VC framework, which means that we work only with the values of (σ_1, σ_2). After some more states which we admit for brevity, this leads to the vanishing of the supremum. We are then left with a sum of two random variables whose sum can be controlled using the Bennett-Bernstein inequality. This is here where the tricky part begins. It is known that averaging over two random variables does not gives you a good control of their expectation. But this can be overcome if we consider many exact replica of the first two variables. Then we have plenty of them, as much as we want! And we can then control the expectation because now the value of m is big. We call this trick, the replica trick. Note the replica trick has been used many times in the invention of algorithm: when exploring the space A of all possible algorithms, the same algorithm has been visited many times but with negligible variances so that if you use an ϵ-insensitive loss functions, these algorithms appear to be equivalent.[4] □

The theorem we just proved should be considered as the dual of the theorem of Vapnik and Chervonenkis. And this should be the case because it is just the dual of it. And we believe this is the more natural setting for you every day design of algorithms. Maybe our approach is complementary to the one of Vladimir and Alexei but we are one step forward because we can infer/compute the probability over the data sets just by looking at the UCI repository database. Just try to do the same with your set of functions and we will talk. Anyway, we insist that our approach shares a lot of common properties with the classical VC framework. One of them is that unfortunately we cannot say much but we try to, or say differently we have our own "no free lunch" theorem but we try to forget it. Here is our no free brunch!! theorem:

[3] Seriously, a small set so that he does not have to do too many experiments.
[4] In fact, embedding nips papers into a vector space, we found a big cluster where all the points were close to each other at a distance less than 0.05 which is the classical significant threshold used in statistics. We ran k-means 50 times but kept coming up with the same single big cluster.



Theorem 2 (No Free Brunch!!) *The generalization error of two datasets for all algorithms is the same:*

$$E_A[R_{gen}{}^{A}[D]] = E_A[R_{gen}{}^{A}[D']]$$

so that there is no better dataset than the one you already have.

The consequences of the theorem are very hard: it means that if you don't do well, then you are not very skilled. We still have not worked out what the researchers had for breakfast (of course, in this remark we retain some hilarity). This leads to the natural consequence that to say anything one should restrict the set of algorithms to some natural set, in a companion paper we prove that the set of datasets restricted to all Bayesian algorithms has infinite dimension. The same is true for the set of all kernel algorithms if you leave the the following free parameters: loss function (hinge loss, pound loss, epsilon-insensitive loss, $\ell_1, \ell_2, \ell_2{}^2, \ell_\infty$, Huber, hyperbolic tangent, Bregman, Beeman, etc.), regularizer (BKHS, $\|w\|^2$, $\sum_i \alpha_i^2$, $\sum_i \alpha_i$, KL divergence, relative entropy, $\log(\max_i \alpha_i)$ etc.) and the set of functions (we do not list here all possible kernels, this will future work to examine the kernel set selection phenomena) and the optimization procedure (perceptron, gradient descent, chunking, quadratic, linear, stochastic, simulated annealing, analytic, random)".
The second major contribution is that our bound has exposed one of the major problems of empirical dataset minimization: the previous de facto method over ten years of nips[5] publications. Clearly, the second term in the bound (which is the confidence interval over the set of datasets) is not taken into account. This leads us to proposing a new construction principle called the Structural Data Sets Maximization (SDSM).

3 Structural Data Sets Minimization

In order to appreciate the following section, we ask the reader for a little patience and to concentrate a little. Assume you have a increasing set of datasets $\mathcal{D}_1, ..., \mathcal{D}_S$ (e.g. USPS, NIST, REUTERS 1, REUTERS 2, ...) which are roughly included in each other:

$$\mathcal{D}_1 \subset \mathcal{D}_2 \subset ... \subset \mathcal{D}_S$$

Then we would like to apply the theorem of the previous section and choose the best datasets for our algorithms (i.e. the one that will be used in the paper presenting these potentially new algorithms). Using the union bound, we found that with probability $1 - \sum_{i=1}^{S} \eta_i$, for all $D \in \cup \mathcal{D}_S$:

$$R_{gen}{}^{A}[D] \leq \min_{D \in \mathcal{D}_i} R_{emp}{}^{A}(D) + O\left(\sqrt{\frac{\Phi(\mathcal{D}_i)}{m}\log(1/\eta_i)}\right) \quad (2)$$

So that we can pick the i^* such that $\Phi(i)$ is minimized over i and choose then the best data set on $\mathcal{D}_{i^*}$. The consistency of this procedure can be ensured by a theorem.
This section means that to find the best data set for your algorithm, you can consider many datasets and compute their capacity and then pick the best one not only in terms of empirical errors (which is strangely often called test error in many papers) but also in terms of this capacity.
Here is a nice picture (Figure 1).

4 Algorithm Ordering Machine

We have now prescribed a formal method for choosing between real datasets. However, we don't really know the capacity of these real world datasets, only approximations can be calculated (although empirical observations show that most published algorithms do well, so the capacity must be quite high). For that reason, we suggest to use toy problems. It turns out that we can prescribe an efficient algorithm for searching for the best toy problem for a given algorithm A.

[5] We do not cite the individual papers, we refer the interested reader to the NIPS volumes.
[6] www.nips.com

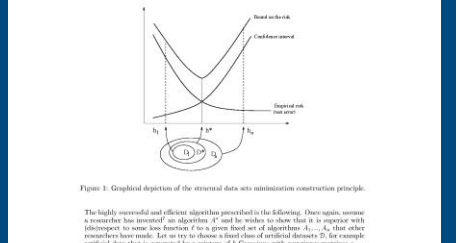


Figure 1: Graphical depiction of the structural data sets minimization construction principle.

The highly successful and efficient algorithm prescribed is the following. Once again, assume a researcher has invented[7] an algorithm A^* and he wishes to show that it is superior with (die)respect to some loss function ℓ to a given fixed set of algorithms $A_1, ..., A_n$ that other researchers have made. Let us try to choose a fixed class of artificial datasets $\mathcal{D}$, for example artificial data that is generated by a mixture of k Gaussians with covariance matrices σ_{ij} in a d dimensional space with a noise model ν and f extra noisy features. Let us define $w = (\nu, d, c, f)$, we have the following theorem:

Theorem 3 *The capacity $\Phi(\mathcal{D})$ is bounded by:*

$$\Phi(\mathcal{D}) \leq \min(R^2\|w\|^2, n)$$

where n is the number of parameters (sum of dimensions of the vectors c, d, c and f) and R is the largest possible value of any coordinates of any vectors in the data sets.

Some people would argue that this theorem is valid only for data sets parameterized by gaussian distribution, but actually it can model a lot of real life problems. The interested reader can read the literature about gaussian processes.
Now, the task is to optimize those parameters such that the algorithm A appears much better than the other ones. Let us choose A=Breno, A_1=SVM, A_2=MARS, A_3=k-NN, $A_4 \cup$=CL5 and we embed this set of algorithms with a uniform distribution to ensure no bias. This can be done with the following minimization:

$$\min_{c,d,e,f} \Phi(\mathcal{D}, d, c, f)$$

subject to: $\quad R_{test}{}^{A}[D] < R_{emp}{}^{A_i}[D] - \epsilon, \ i = 1, .., 4$

which is equivalent to:

$$\min_{w=(\nu,d,c,f)} \|w\|^2$$

[7] i.e, slightly altered nonsense else's.

Algorithm	Without noise	With noise
Linear SVM	0.8 (0.005)	0.7 (0.01)
Poly-2 SVM	0.5 (0.1)	0.5 (0.2
Poly-5 SVM	0.7 (0.2)	0.7 (0.1)
Poly-10 SVM	0.9 (0.8)	0.8 (0.6)
Mars	0.7 (0.2)	0.4 (0.1)
Breno	**0.001 (0.00009)**	**0.002 (0.0001)**

Figure 2: Results for $w = (0.00001, 5000354820.34, 3.14159, 2.735, 1, 2, 3, 4, 5, 6, 7, 8, 9, -1, -2, -3, ...)$. w corresponds to the parameters of the generated data set. Unfortunately, we have difficulties interpreting w.

subject to: $\quad R_{test}{}^{A}[D] < R_{emp}{}^{A_i}[D] - \epsilon, \ i = 1, .., 4$

If you wish to find a dataset where your algorithm does not achieve 0% test error you can easily generalize this algorithm to the linearly inseparable case by introducing slack variables ξ_i.
The closeness to SVM is striking. Note that we are maximizing the margin between the algorithms to ensure a paper is accepted. The relation with statistical tests is open and has not been analyzed rigorously yet, but an upcoming work is statistical test selection so that even with small margin you can have a strong result.

5 Experiments

We proceed to prove our methodology by showing we can always find a good solution even for bad algorithms So we proceed with the example given above. We must admit that it has already been done in the literature but we provide a deeper analysis and also better margin. In table 2, we present the results we got with the best value for w.
Note that it is very difficult to discriminate between a linear SVM and a poly-2 SVM, both of them seem to have a similar behavior and we were not able to worsen the results of the poly-2 SVM although it would have been nice. Poly-2 SVM performs well on a large number of data sets so the optimization was difficult, this may explain why we got strange value for w. On the other hand, it is quite clear in the table that Breno is much better than all the other algorithms even much better than MARS although MARS has some common properties with Breno. Thus our algorithm was able to discriminate between ϵ distances in the space of A. We omit our other experiments for brevity but the results were good.

6 Conclusion

We will now reiterate what we said before, we repeat the abstract and introduction. The problem with any unwritten law is that you don't know where to go to erase it. This is the same for other matters by the way. Consider for instance the notations, it is assumed that in σ I always refer to the number of training examples. Sometimes, it is is also but this occurs mainly when the authors are new in the field. Note that we also are new in the field but we did not use n. On the other hand, we have handled quite in a nice way the use of greek letters. Anyway, the question we are discussing right now is to know when and how to step an unwritten law. We believe this could be the place and the time, and, as a mark of courage, we will not reiterate the introduction. This may then sound weird that we, as outsiders, put a stone in the sea of nips paper. We do not know the hydrodynamics laws of such a sea. We do not who discovered water but we're pretty sure that it wasn't a fish. Not even a big fish with a latin name.
Let us stop being polemic for a while and come back to our contribution. Central to our new research is the idea to improve the presentation of algorithms in literature and to make them more appealing. We defined a new notion of capacity for data sets and derived a methodology. The experiments showed that even for not so good algorithms, you can show that they are significantly better than all the other ones. The message is strong and may not be understood at a first reading so we insist here to avoid any confusion: we emplore all researchers to dig out their old failed algorithms and turn them into successful ones.

To be complete, we present in this last paragraph the future works we plan to do one day. We will make the link between dataset selection and the human neural information processing which so far researchers have shown happens, in females, in the human brain[8]. We will consider whether dataset selection is implemented via chemical stimulation of the neurons, possibly in the hippocampus. In humans it could consist of, when failing to learn a task, buggering off and learning something else instead[9].
At last, we would like to say a mild word of caution. We hope that the community learns from this break-through and applies our methodology in their future research or they will get left behind: our algorithms will far outperform theirs.[10]

References

[1] R. Shapire, Y. Freund, P. Bartlett, and W.S. Lee. *Boosting the margin: A new explanation for the effectiveness of U.S voting methods. The Anals of Statistics*, 1998.
[2] N. Cristianini and J. Shawe-Taylor. *Data Set Selection for Dummies*. Eghum University Press, 2005.
[3] P. Bartlett. *The Sample Complexity of Pattern Classification with Neural Networks: the Number of Catalonas is More Important than the Number of Readers. Brouhaf Transactions* , 1998.
[4] D. MacKay. *I did it my way.* . *Self-published*, 2002.
[5] G. Fung and O. L. Mangasarian. Data Selection for Support Vector Machine Classifiers. Proceedings of KDD'2000, 2000.
[6] J. Friedman; T. Hastie; and R. Tibshirani. *Data sets for additive models: a statistical view of bragging and boasting*. Technical report, Technical Report, Department of Statistics, Stanford University., 1998.
[7] C.M. Bishop. *Neural Net works for Pattern Recognition*. Oxford University Press, 1995
[8] D.H. Wolpert. *The lack of a priori distinctions between learning algorithm research*, Neural Computation, 1996.
[9] V.N. Vapnik and A.Y Chervonenkis and S. Barnhill. *The VC Wing: Investment Secrets from the Wizards of Venture Capital*, Bicroiff Publishing (now out of print) , 2001.

[8] Note that for ankles, smart people conjectures it should be in some unknown other place, others yet conjecture it doesn't exist at all. We will refer to this as the D-spot and note that at the least it is very hard to find, at least if the guy is fat.
[9] Notice how many people learn to juggle around exam time .
[10] Finding the occasional straw of truth amoth in a great ocean of confusion and handsome empires intelligence, vigilance, dedication and courage. But if we don't practice these tough habits of thought, we cannot hope to solve the truly serious problems that face us – and we risk borrowing a notion of suckers, up for grabs by the next charlatan who comes along.

UCMERCED

Data Set Selection

Data Set Selection

Doudou LaLoudouana* **and Mambobo Bonouliqui Tarare**
Lupano Tecallonou Center
Selacie, GUANA
doudoula3@hotmail.com, fuzzybear@yahoo.com

Abstract

We introduce the community to a new construction principle whose practical implications are very broad. Central to this research is the idea to improve the presentation of algorithms in the literature and to make them more appealing. We define a new notion of capacity for data sets and derive a methodology for selecting from them. The experiments show that even for not so good algorithms, you can show that they are significantly better than all the others. We give some experimental results, which are very promising.

Data Set Selection

References

[1] R. Shapire, Y. Freund, P. Bartlett, and W.S. Lee. *Bushing the margin: A new explanation for the effectiveness of U.S voting methods. The Anals of Statistics*, 1998.

[2] N. Cristianini and J. Shawe-Taylor. *Data Set Selection for Dummies.* Egham University Press, 2005.

[3] P. Bartlett. *The Sample Complexity of Pattern Classification with Neural Networks: the Number of Citations is More Important than the Number of Readers. Biowulf Transactions* , 1998.

[4] D. MacKay. *I did it my way. Self published*, 2002.

[5] G. Fung and O. L. Mangasarian. Data Selection for Support Vector Machine Classifiers. Proceedings of KDD'2000, 2000.

[6] J. Friedman, T. Hastie, and R. Tibshirani. *Data sets for additive models: a statistical view of bragging and boasting.* Technical report, Technical Report, Department of Statistics, Stanford University., 1998.

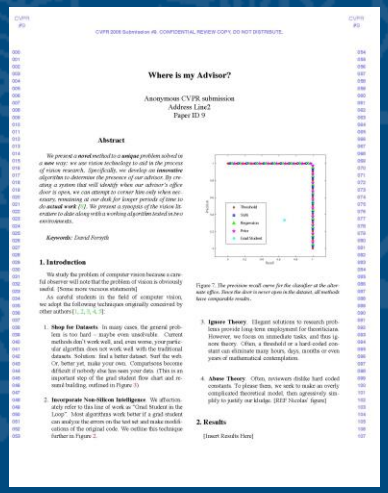
[7] C.M. Bishop. *Neural Not works for Pattern Recognition*, Oxford University Press, 1995

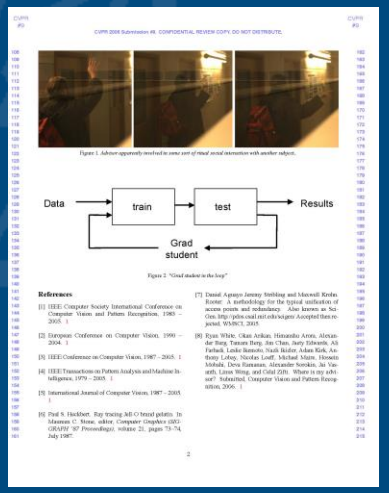
[8] D.H. Wolpert, *The lack of a priori distinctions between learning algorithm research*, Neutral Computation, 1996.

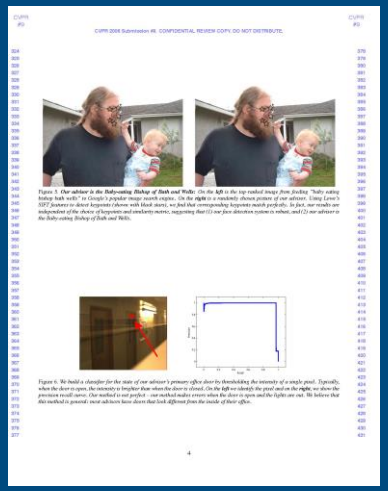
[9] V.N. Vapnik and A.Y Chervonenkis and S. Barnhill, *The VC Way: Investment Secrets from the Wizards of Venture Capital*, Biowulf Publishing (now out of print) , 2001.

(originally) [6] ... a egotistical view of bragging and boasting.....

Where Is My Advisor?





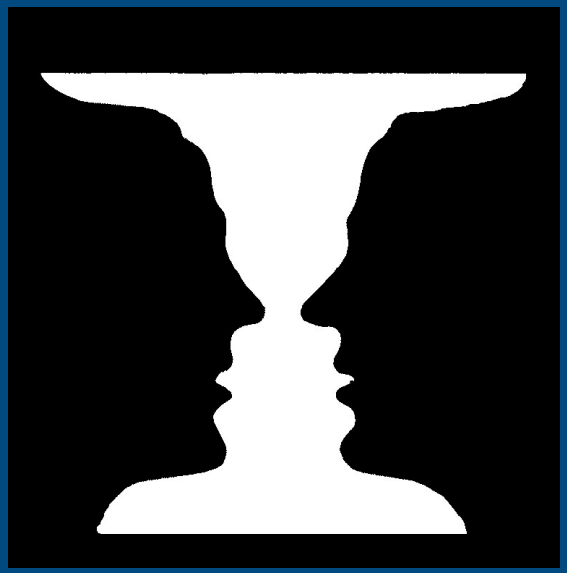


UCMERCED

Ask Someone to Proofread

- Certainly your advisor
- Polish your work
- My story

Paper Gestalt



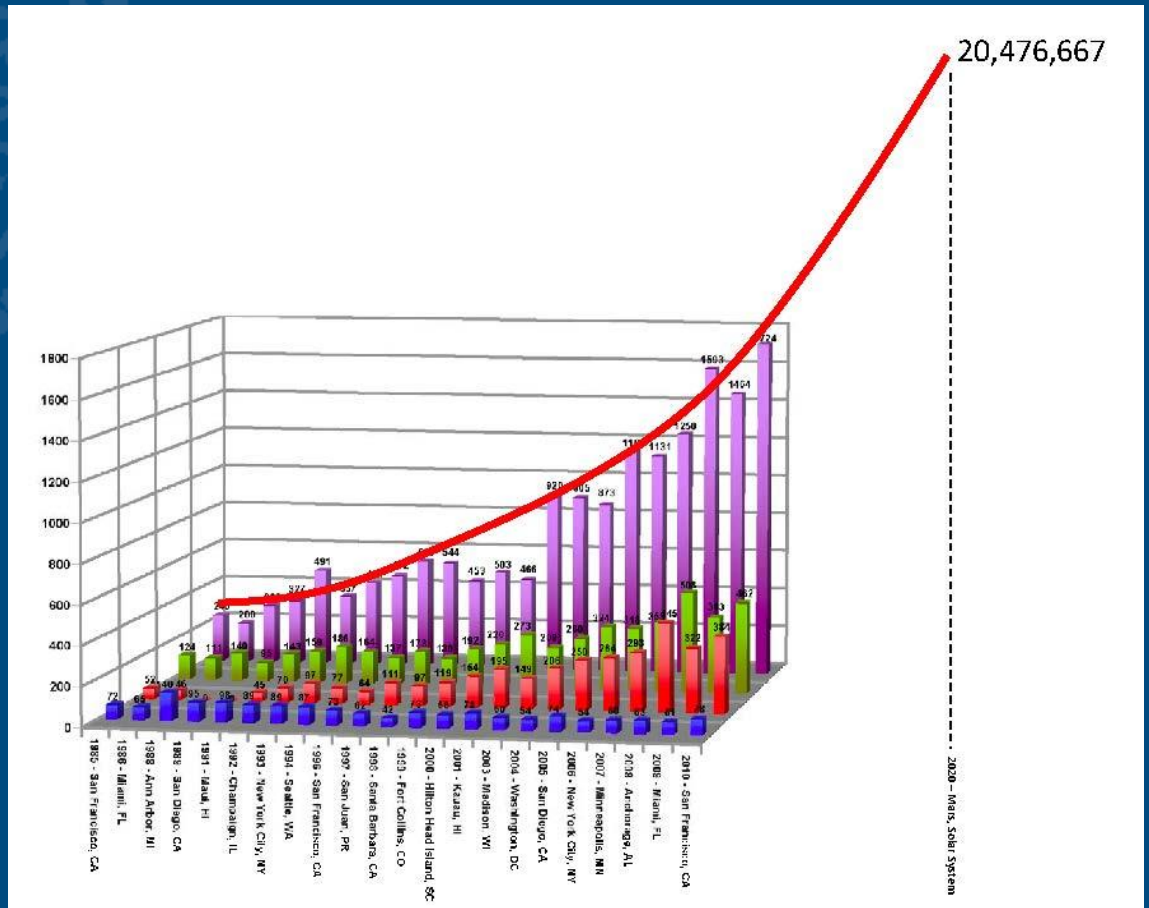


Figure 1. **Paper submission trends.** The number of submitted papers to CVPR, and other top tier computer vision conferences, is growing at an alarming rate. In this paper we propose an automated method of rejected sub-par papers, thereby reducing the burden on reviewers.

Paper Gestalt

- CVPR 10 by Carven von Bearnensquash, Department of Computer Science, University of Phoenix
- Main Point: Get your paper looking pretty with right mix of equations, tables and figures

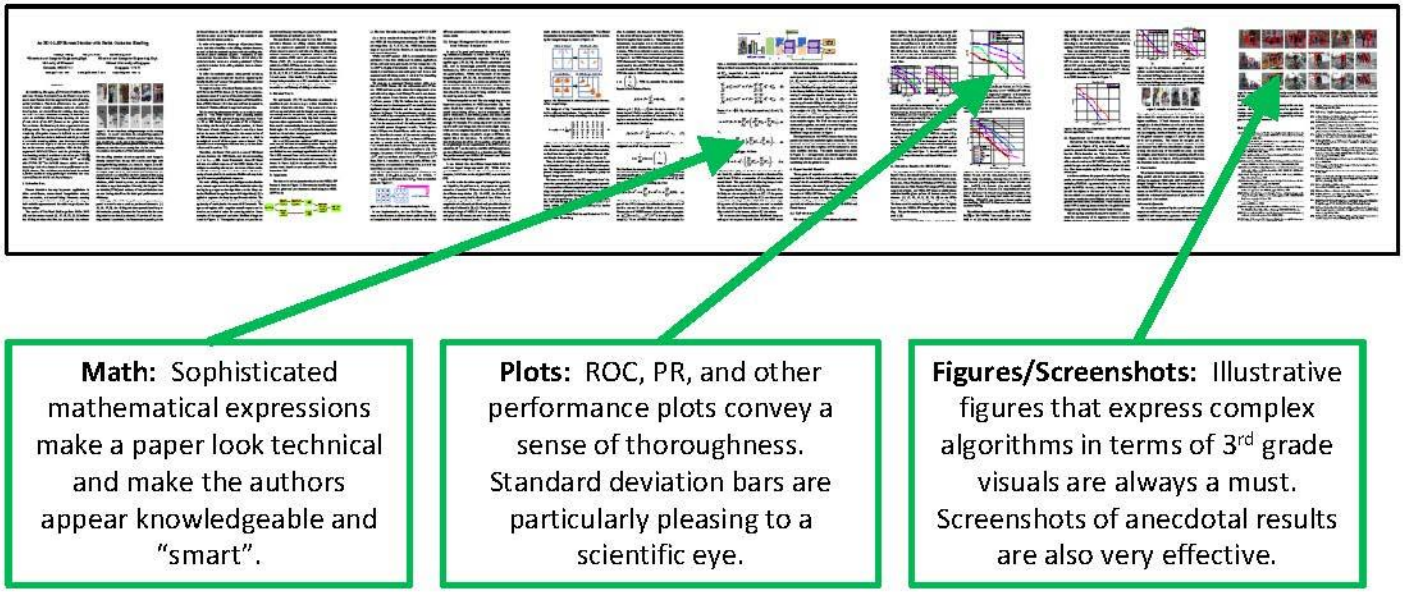


Figure 6. **Characteristics of a "Good" paper.**

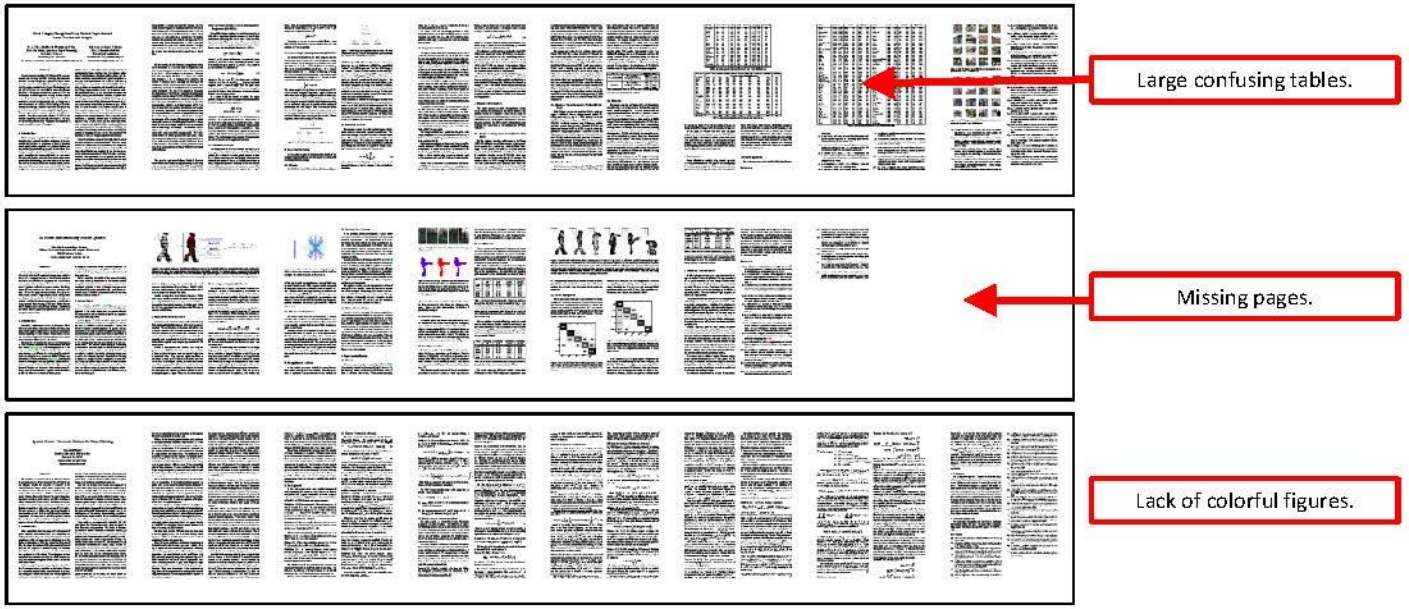


Figure 7. **Characteristics of a "Bad" paper.**

Tools

- Google scholar h-index
- Software: publish or perish
- DBLP
- Mathematics genealogy

- Disclaimer:
  - h index = significance?
  - # of citation = significance?

Basic Rules

- Use LaTeX
- Read authors' guideline
- Read reviewers' guideline
- Print out your paper – what you see may NOT be what you get
- Submit paper right before deadline
  - Risky
  - Exhausting
  - Murphy's law
- Do not count on extension

UCMERCED

Lessons

- Several influential papers have been rejected once or twice
- Some best papers make little impact
- Never give up in the process

Karma?

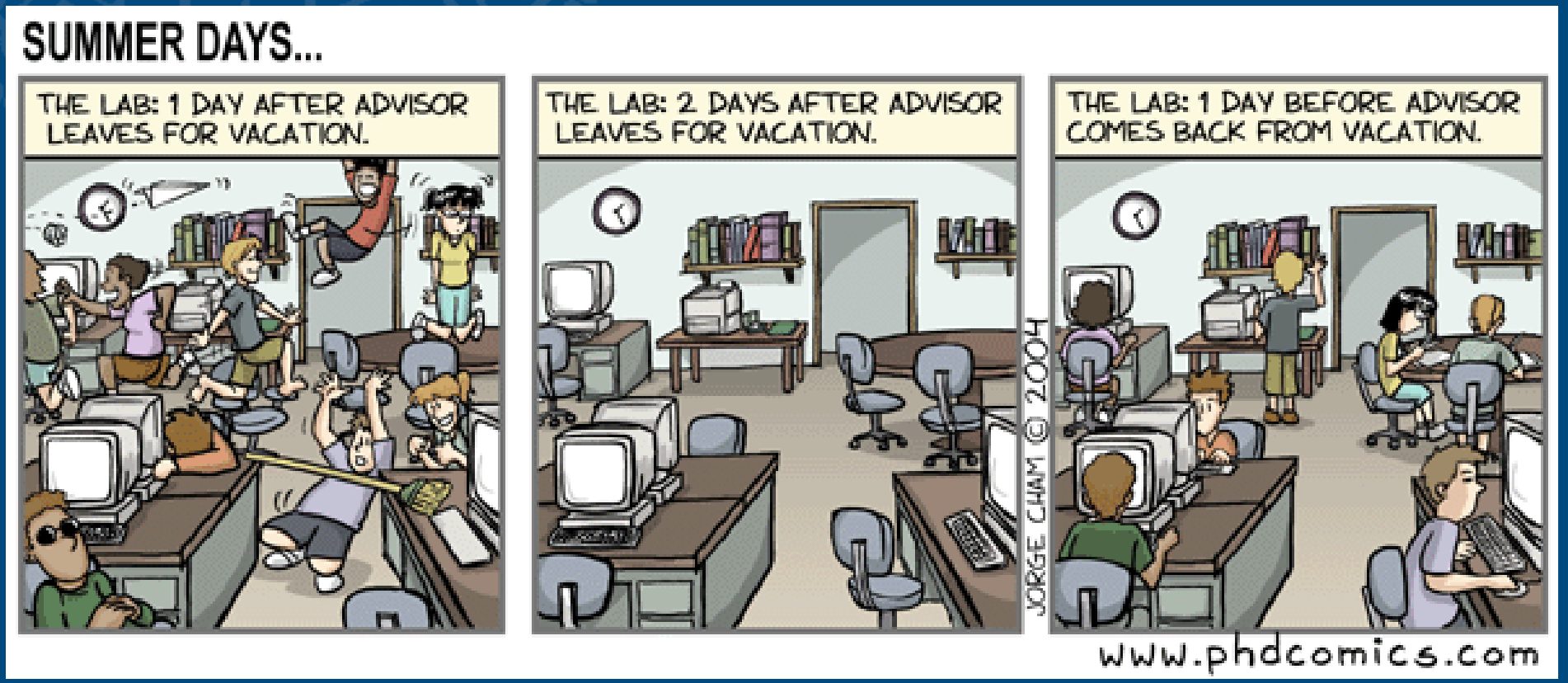
WHAT YOU THINK OF YOUR PROFESSOR vs. TIME
SHE OR HE IS A GENIUS!!
OK, THEY'RE ONE OF THE BEST, BUT MAYBE NOT THE BEST.
THEY'RE "OK." I COULD DO BETTER.
DO THEY EVEN KNOW WHAT THEY'RE DOING??
AT LEAST THEY CAME TO MY THESIS DEFENSE.
NICE GAL OR GUY.
1st YEAR
2nd YEAR
3rd YEAR
4th YEAR
5th YEAR
TIME
JORGE CHAM © 2014
WWW.PHDCOMICS.COM

Your Advisor and You

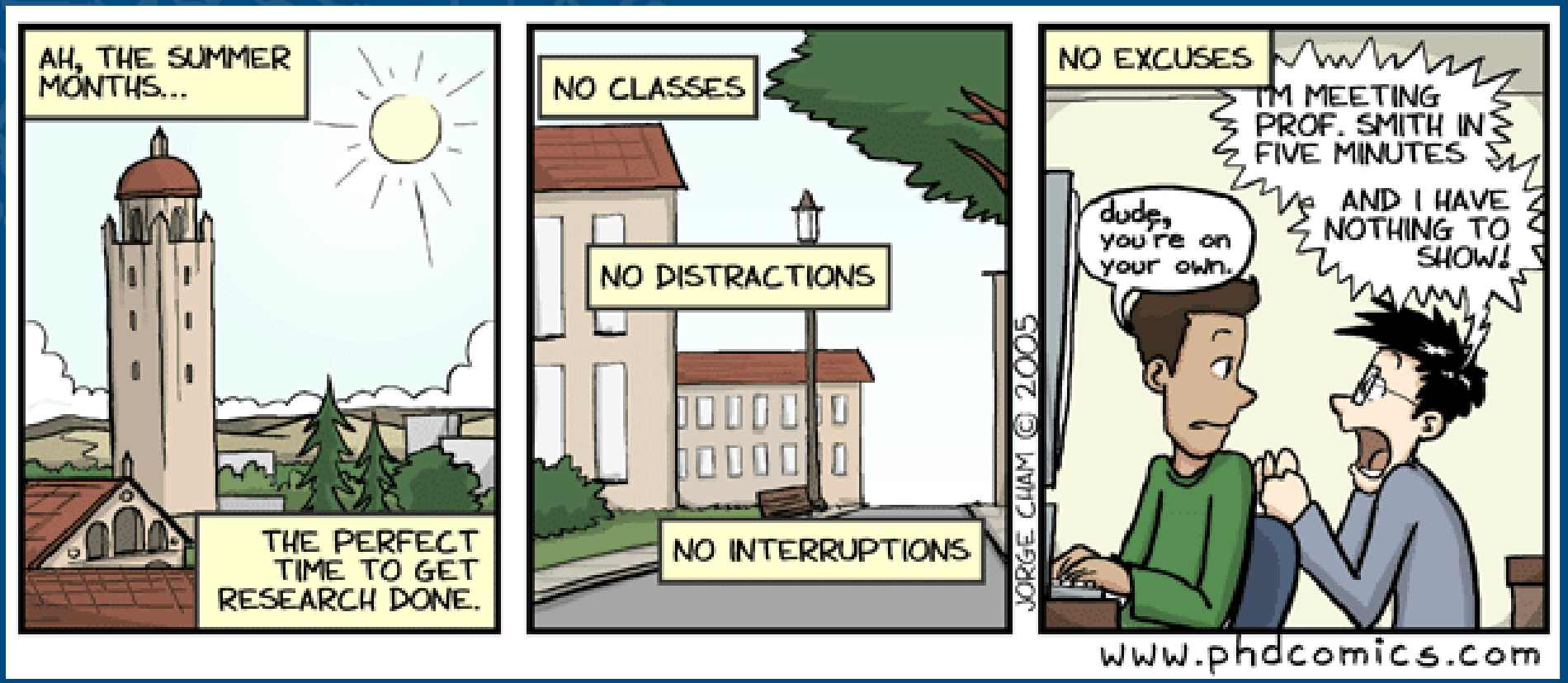
- Suggesting a research topic
- When your advisor presents your work
- When you explain your work
- Demos
- Good results

Start Working Early!

- Write, write, write…
- Ask others for comments



Work Hard in the Summer



Quotes from Steve Jobs

- “ I'm convinced that about half of what separates successful entrepreneurs from the non-successful ones is pure perseverance. ”

- “ Creativity is just connecting things. When you ask creative people how they did something, they feel a little guilty because they didn't really do it, they just saw something. It seemed obvious to them after a while. ”