Effect of Signal Sparsity, Signal Density and Noise Level on Rotational Equivariant Features for Image Classification

Aliraza Punjani   Apoorv Kulshreshtha   Jamshed Shapoorjee   Plaban Mohanty

{ amp2280, ak3963, js4962, pm2878 }@columbia.edu

Abstract

Current convolution methods for feature learning are already translation equivariant i.e. translation in input image produces a proportionate translation in feature maps. However, this is not true for rotational equivariance. A lot of recent research has been focussed on ensuring rotational equivariance for the same. Our research focusses on characterizing effect of different input image parameters, like sparsity, density and noise, on effectiveness of the learnt rotational equivariant features proposed in the recent work of Harmonic Convolutions. Feature maps learnt using Harmonic Convolutions exhibit equivariance to patch-wise translation and 360 degree rotation. These variant of normal convolutions use parameter-efficient and low computational complexity representation, thereby encoding complicated rotational equivariance within the network. In this paper, we show the effectiveness of rotational equivariance features for image classification as the sparsity, density and noise levels of the input image vary.

1. Introduction

Traditional convolutional networks are the one of the popular methods for Image Classification. Due to various advancements over the years and the availability of more computational power convolutional neural networks have been performing exceptionally well. Although the traditional CNNs are translational invariant i.e. they can correctly classify an image even if the image is translated along the X or Y axis they are still cant perform well on images that are rotated. Since in real-world scenarios the images are taken from different angles and they are rotated along different axis it is important to be able to detect images which are rotated. Recent research suggests using harmonic convolutions to learn rotational equivariant features. Our paper explores how these harmonic convolutions work and their performance on rotated images. Another issue in classifying images is that the images which are used in various applications have issues like low quality and some unavoidable noise. Hence it is important that if we are using harmonic convolutions they should be able to recognize these images and should not be affected with factors like these. Our research focuses on how to get the best results on rotated images and how factors like noise, image density and image sparsity affect these methods.

2. Related Work

In a large number of applications, the image data used will have images that are from different angle and rotated along many axis. The paper by Worrall et al. [1], on harmonic Network describes how harmonic convolution based filters can be used for learning rotational equivariant features. We use this approach to study how these harmonic features learnt are affected by various parameters like noise, sparsity, and loss of quality during image compression. Various papers like Dodge and Karam[2] have explored how Image Quality and how it affects the ability of a Neural Network to accurately recognize an image. They have pointed out some interesting conclusions about how the networks are susceptible to factors like noise and blur but are resilient to compressions. Basu et al.[4] has explored the MNSIT dataset with additional Gaussian noise and proposed some changes to the architecture since the traditional ones did that perform well. Rigamonti et al. [3] describes the effect of sparse representation of images on image classification he concludes that at a reasonable noise level sparsity does not affect the classification of images and given the high computation cost in sparse encoding the images it can be avoided. Zabala et al.[4] in their paper in compare the affect of JPEG and JPEG-2000 compression on Image classification and concluded that the type of compression used also plays an important role in classifying images.

In this paper, we aim to bring out the differences in how individual factors in an image affect a traditional CNN differ form a neural network trained using harmonic convolutions. We alter the images by introducing gaussian noise, sparsity and change of quality to see how these factors affect the rotational equivariant features learnt using harmonic convolutions. We use the CIFAR-10 dataset to run

our experiments. This dataset contains 60000 32x32 color images in 10 classes, with 6000 images per class

3. Problem Analysis

Harmonic Convolutions hard bake 360 degree rotational equivariance into their feature representations by restricting the convolution filters to be from the circular harmonics family. In the following sections, we will discuss the properties of circular harmonics which help the network learn rotation equivariant features. To reiterate, rotational equivariance implies that a particular rotation in input image produces a proportionate rotation in feature maps.

3.1. Circular Harmonics Equivariance

We describe an image using polar co-ordinates r and ϕ as F(r , ϕ). This can be understood better by looking at Figure 1. Please note that this is just to depict the polar co-ordinates of a filter. In actual setting, the filter will be applied patch-wise to the image and therefore, a single image patch will be described using this method.

We will now show that there exists a filter W_m such that the cross-correlation of F with W_m yields a rotationally equivariant feature map. This condition is satisfied when W_m is a circular harmonic of the form W_m = R(r) $e^{i(m\phi+\beta)}$ for some m belonging to integers. Consider the rotation of original image by θ which leads to a new image F(r, $\phi - \theta$). The cross-correlation for the rotated image is

$$[W * F(r, \phi - \theta)] = \int W(r, \phi) F(r, \phi - \theta) \, dr \, d\phi$$
$$= \int W(r, \phi' + \theta) F(r, \phi') \, dr \, d\phi'$$

where $\phi' = \phi - \theta$. If we replace W_m to be of the form R(r) $e^{i(m\phi+\beta)}$, then the integral transforms as:

$$[W * F(r, \phi - \theta)] = \int R(r) e^{i(m(\phi'+\theta)+\beta)} F(r, \phi') \, dr \, d\phi'$$
$$= e^{im\theta} \int R(r) e^{i(m\phi'+\beta)} F(r, \phi') \, dr \, d\phi'$$

When rotation $\theta = 0$, then $\phi = \phi'$. Therefore, the above equation can be written as

$$= e^{im\theta} \int R(r) e^{i(m\phi+\beta)} F(r, \phi) \, dr \, d\phi$$
$$= e^{im\theta} [W * F(r, \phi)]$$

Hence, we observe that the cross-correlation of the rotated signal F(r, $\phi - \theta$) with harmonic filter W_m = R(r) $e^{i(m\phi+\beta)}$ is equal to the response at 0 rotation [W * F(r, ϕ)] , multiplied by a complex phase shift $e^{im\theta}$. Thus, we have shown that cross-correlation with W_m yields a rotationally equivariant feature mapping when W_m is a circular harmonic.

Figure 1. An image describing polar co-ordinates with image center as origin

3.2. Chain rule for cross-correlation of circular harmonics

In this section, we will show that the rotation order of a feature map, that we obtain after subsequent cross-correlations in each layer, is equal to the sum of the rotation orders of the filters in the chain. Let us consider cross-correlation of a θ rotated image F(r, $\phi - \theta$) with a filter W_m and then followed by another cross-correlation with W_n. As shown previously, we write the response of first cross-correlation as $e^{im\theta}$[W * F(r, ϕ)] . Then we can write the chain rule of harmonic cross-correlation as:

$$W_n * [W_m * F(r, \phi - \theta)] = W_n * e^{im\theta}[W_m * F(r, \phi)]$$
$$= e^{im\theta}(W_n * [W_m * F(r, \phi)])$$
$$= e^{im\theta}e^{in\theta} [W_n * [W_m * F]]$$
$$= e^{i(m+n)\theta} [W_n * [W_m * F]]$$

Therefore, we see that chained cross-correlation results in a summation of the rotation orders of the individual filters W_m and W_n.

3.3. Circular harmonic filter architecture

As we described in the previous sections, a Harmonic filter is defined as W_m = R(r) $e^{i(m\phi+\beta)}$, so that it can capture the rotation equivariance while image classification. We also showed that whenever a test image is rotated, the cross-correlation of the rotated image with the learned filter, results in a product of $e^{im\theta}$ with [W * F(r, ϕ)], where
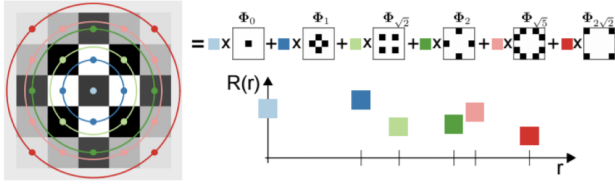
Figure 2. Filter Architecture showing mapping of concentric harmonic filter profile to the cells of a square filter

[W * F(r, ϕ)] is the original non-rotated test image. Hence, the classifier is able to learn that any $e^{im\theta}$ multiple of the original cross-correlation is the same image rotated by θ, and thereby classify it correctly. Therefore, this clearly defined mapping helps achieve rotation equivariance of filters, which in turn leads to global rotation invariance of the network in predicting rotated images. This is the reason, the network is able to classify rotated images with equally good accuracy.

We will now delve deeper into the architecture of a Harmonic filter. The learnable parameters of a harmonic filter $W_m = R(r)\ e^{i(m\phi+\beta)}$, are the radial profile R(r), which is a function of the radius r from the origin, and the per-filter phase offset β. For a n*n filter, the number of radial profile elements is equal to the number of rings of equal distance from the center of the filter. An example of this filter can be seen in Figure 2. This filter is a 5 * 5 filter with 6 rings of equal distance from the center of the filter. The smallest ring is just a single point. It is called a 5*5 filter because we are mapping the radial profile to the elements of a square filter with has length 5 and hence dimensions 5*5. And hence, this filter has 6 radial profile terms and 1 phase offset to learn. This is in contrast to a regular convolution filter which has 25 parameters to learn (because of 5*5 dimension of the filter).

Now, that we have explained the architecture of the harmonic filters, there is one very important characteristic to notice. Because we are learning a complete radial profile which depends only on the radius, the filter values for a particular radial profile are same, with only difference in sign. This can be perceived as an advantage as well as a disadvantage. Advantage because this kind of constraint enables the network to learn a rotation invariant classifier. And a disadvantage because the constraint doesn't let the network learn the filter freely during training , so that it can perform the best. In the next sections, we will see how this constraining of filters serves as an advantage as well as a disadvantage when looking at different applications.

To reiterate, till now the paper gives the reader a background of how Harmonic Networks work and how they are able to hard-bake rotation equivariance. More refer-ence on this can be found in the paper [GIVE REFERENCE HERE]. From now on, we will discuss our main focus which describes how change in signal sparsity, signal density and noise level affects rotational equivariant features.

4. Parameters Affecting Rotational Equivariance

4.1. Noise

Most of the image data used in real world applications are noisy in nature. One of the most important applications of image recognition and classification falls under the domain of supervision, intelligence and surveillance and despite the massive leaps in vision technology, the images captured by devices are quite noisy. While there has been good research in the fields of de-noising these images , the effort and computation required outweighs the performance of such methods. Hence, it is quite essential to analyze the effect of noise and its variations on harmonic filters to see if it still retains its accuracy and rotation equivariance . We also framed analytical conclusions based on the theory behind harmonic filters and performed experiments to confirm our understanding.

Types of Noise :

An image could have multiple variations of noise incorporated into it. Some of the most common ones are :

1. *Gaussian Noise* :

This is a form of additive noise added independently at each pixel. The added value is independent of signal intensity and is based on a gaussian distribution. One of the major reasons for this form of noise is the thermal noise (Johnson Nyquist Noise). This is one of the most common form of noise that arises due to low background illumination , increased temperature etc.

2. *Shot Noise* :

This is a noise caused due to exposure fluctuations during the time the image is captured. It is caused different counts of photons sensed at a exposure level. It is dependent on the image intensity and is based on a Poisson distribution. As a result, apart from lower values of intensity its effects mimic Gaussian distribution

3. *Uniform Noise* :

It is caused by quantizing pixels of an image to discrete levels .the distribution of such a noise is uniform.

4. *Periodic Noise* :

This noise appears as a periodic or iterating pattern over the image and is usually caused by electromagnetic inter-

ference while capturing the image.

Analysis of effect of noise :

For our analysis, let us assume Gaussian noise . As it is a form of additive noise that is independent of intensity , the variation is induced by the general std deviation of the gaussian distribution the noise is based on . Random noise values are added to the pixels from within the gaussian distribution.

As we start off with lower values of std deviation , the variation in the values of noise added to the pixels are not significantly large. Hence the original image can still be retrieved from the noisy image as the added noise value is nearly equal with less variation across all the pixels.

Considering our equation no (TBA), upon normalization, the harmonic filters are still able to get a multiple of the original cross correlation from the cross correlation of the noisy rotated image. The hard baked harmonic filters identify the corresponding pixel values which although modified by a similar value throughout , are recognizable as their variation is less and hence they still retain their original pattern in the patch even after being scaled up (or down, depending on the noise added). Hence the harmonic filters perform almost as well with noise with low variance as they do without noise.

However, as the variance or std deviation of the gaussian distribution to which the noise belongs to increases , the variation of randomness in the addition of values to the pixels increases . As a result the pattern in the patch that we described earlier is lost i.e the change in pixel values are not constant and vary a lot. This makes it highly difficult for the harmonic filters to retrieve the cross correlation as a multiple of the original cross correlation. The F(r,phi) that we get here might be different from the original noiseless image .Hence it is difficult to classify the representation of the image. Although harmonic convolutions still preserve rotational equivariance, the original pattern of the image is altered and as a result the harmonic filters are not able to classify the image properly leading to lower accuracy.

This property makes the harmonic convolution exhibit similar traits as convolutional filters and hence the results reflect a similar pattern.

Experimental Setup :

To observe the variation in performance of the filters with noise, we chose to experiment with adding varying levels of Gaussian Noise. We used python libraries to introduce both noise and rotation in the images. To produce variation in noise levels, we kept the mean as 0 and increased the standard deviations or variance to observe the effects.

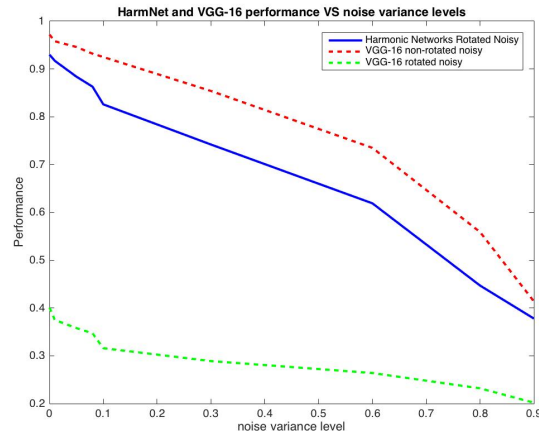To provide a basis of comparison for such results, we also

Figure 3. Graph showing trend of accuracy with increase in noise

ran a similar setup with convolutional neural networks . To further study only the effects of noise on harmonic filters and factoring out the rotational equivariance property , we also ran the experiments with non rotated noisy data on CNNs. The neural network architecture used for both the CNNs was VGG-16 .¡Ref¿ The entire experimental analysis was done on CIFAR-10 image dataset with 60k images

Experimental Analysis :

As can be observed from the graph in Fig 3 , the harmonic filters perform well in the lower values of variance in additive noise but steadily decrease as the variance increases .The same trend is observed in neural networks with non-rotated noisy images as well .An interesting point to observe is that the accuracy of harmonic filters is less by a near constant amount with respect to CNNs on non-rotated image across the entire graph. It has to be remembered that while the harmonic convolutions are on the rotated images, this CNN is on non-rotated images. The decreased performance is due to the restricted condition of the harmonic filters which learn parameters for rotation equivariance, however filters for CNN are free to learn and are unrestricted. With same kind of data, both rotated and noisy, CNNs perform extremely poor in comparison to harmonic filters

4.2. Sparsity

Sparse representations of image data are widely used and recognized in the field of vision and classification experiments. Their increased popularity owing to their suitability as a regularizer for general inverse problems as well as their ease in supportability in high dimensional spaces have made them widely used in almost all image representation problems. The ease in storing and accessing sparse representations also make it highly convenient for their use. With current research showcasing increased performance

in certain cases of classification with sparse representations , it is highly essential that we analyze its effect on the harmonic filters to observe the change in rotation equivariance properties.

Ways to induce sparsity in images :

Sparse approximation problem is a highly researched topic with efficient algorithms to solve this problem in place . We list some of the most commonly used algorithms :

1. *Projected Gradient Descent* :

This is similar to normal gradient descent where the gradient shows the direction for movement. Since we are looking for a sparse solution, the putative solutions are projected onto the sparse scaffold of k vectors.

2. *LASSO* :

This regression analysis method is widely used along with its multiple variations. It solves the L1 norm version of the sparse approximation method . Fused Lasso ,Group Lasso and other variations are used for generating sparsity.

3. *Proximal Methods* :

Proximal methods including proximal gradient descent is also used for induced sparsity in images .With thresholding /hard thresholding , the sparsity induced by this method forms the basis for many classification experiments.

Analysis of Sparsity on images :

To understand the effect of sparsity, we need to refer to the hard baked filters that are applied on patches of data. These filters as explained in the section above are concentric filters applies on pixels that in turn produce rotation equivariance. This method relies on the pattern of pixels on which the filter is applied. Now as we induce sparsity, some or most of the pixels are reduced to 0. Now, even though the pixels might not be relevant for image classification , this hinders the basic property of the pattern in the patches that the harmonic filters rely on. Hence the hard baked filters which learn in a constrained manner are not able to produce rotation equivariant features.

We can visualize this with reference to fig. where the harmonic paths are applied on a patch. Now , if we decrease the values of pixels to 0 , the purpose of the harmonic filter is defeated as with some removed pixel values , the filters do not apply on the similar kind of images as they have been trained on. As a result, in terms of cross correlation , the cross correlation value of the rotated sparse image is not a multiple of the cross correlation of the original image. Hence , the classification isnt accurate.

As this adversely affects the filters, the increase in sparsity should cause rapid decrease in performance of the classifier. Now, it can be observed that since these constraints
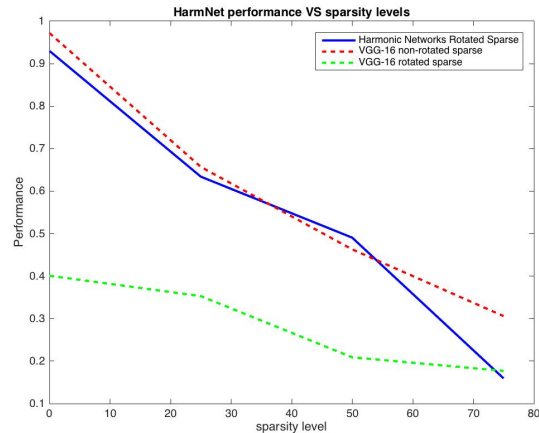

Figure 4. Graph showing trend of sparsity with increase in noise

are not present in the normal CNN filters , this massive decrease in performance is not observed in the case of normal convolutions. The constrained filters which produce rotation equivariance are also the cause for the decrease in accuracy of classification for harmonic convolutions .

Experimental setup:

The setup is the same as the experiments with noise. The variation in sparsity is produced by Python libraries and the percentage of sparsity is increased to observe the effect. The image data is CIFAR-10 with 60k images. Similar to experiments with noise, we have kept two standard observations for comparison CNN with non-rotated sparse images and one with rotated sparse representations.

Experimental Results:

As expected, there is a steep decline in the performance of the harmonic filters on increase in sparsity. With higher levels of sparsity, the filters perform poorly and lose their rotation equivariance property. Further on extremely high values of sparsity it is observed that due to the constrained filters in harmonic networks, the accuracy falls even below CNN on rotated ,sparse images. This stands as a testimony to the adverse effect of sparsity in harmonic networks.

5. Conclusion

6. Future Work

Our research clearly points out how various factors of an image like image sparsity, noise level and quality affect traditional convolutional networks in ways that do not hold true for the rotation equivariant networks. All the factors tested in this paper make sure that the images space does not change, we would like to work on how lower dimensional feature representations like compressed sensing and

sketching affect harmoncic convolutions. Lower dimensional representations can also speed up the time taken to train the network and save on essential resources. Hence it would be interesting to see the tradeoffs between the use of lower dimensional representations for learning rotational equivariant features and the loss of accuracy

7. References

[1] Worrall, D. E., Garbin, S. J., Turmukhambetov, D., and Brostow, G. J. (2016). Harmonic Networks: Deep Translation and Rotation Equivariance. arXiv preprint arXiv:1612.04642.

[2] Dodge, S., and Karam, L. (2016, June). Understanding how image quality affects deep neural networks. In Quality of Multimedia Experience (QoMEX), 2016 Eighth International Conference on (pp. 1-6). IEEE.

[3] Rigamonti, R., Brown, M. A., and Lepetit, V. (2011, June). Are sparse representations really relevant for image classification?. In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on (pp. 1545-1552). IEEE.

[4] Zabala, A., Pons, X., Daz-Delgado, R., Garca, F., Aul-Llins, F., and Serra-Sagrist, J. (2006, July). Effects of JPEG and JPEG2000 lossy compression on remote sensing image classification for mapping crops and forest areas. In Geoscience and Remote Sensing Symposium, 2006. IGARSS 2006. IEEE International Conference on (pp. 790-793). IEEE.

7.1. Submitting Final Camera-Ready Copy

The final versions of papers accepted for publication should follow the same format and naming convention as initial submissions, except of course that the normal author information (names and affiliations) should be given. See Section 8.3.2 for details of how to format this.

The footnote, "Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute." must be modified to "*Proceedings of the 34th International Conference on Machine Learning*, Sydney, Australia, 2017. JMLR: W&CP. Copyright 2017 by the author(s)."

For those using the LaTeX style file, simply change `\usepackage{icml2017}` to

$$\text{\usepackage[accepted]{icml2017}}$$

Authors using **Word** must edit the footnote on the first page of the document themselves.

Camera-ready copies should have the title of the paper as running head on each page except the first one. The running title consists of a single line centered above a horizontal rule which is 1 point thick. The running head should be centered, bold and in 9 point type. The rule should be 10 points above the main text. For those using the LaTeX style file, the original title is automatically set as running head using the `fancyhdr` package which is included in the ICML 2017 style file package. In case that the original title exceeds the size restrictions, a shorter form can be supplied by using

```
\icmltitlerunning{...}
```

just before `\begin{document}`. Authors using **Word** must edit the header of the document themselves.

8. Format of the Paper

All submissions must follow the same format to ensure the printer can reproduce them without problems and to let readers more easily find the information that they desire.

8.1. Length and Dimensions

Papers must not exceed eight (8) pages, including all figures, tables, and appendices, but excluding references and acknowledgements. When references and acknowledgements are included, the paper must not exceed ten (10) pages. Acknowledgements should be limited to grants and people who contributed to the paper. Any submission that exceeds this page limit or that diverges significantly from the format specified herein will be rejected without review.

The text of the paper should be formatted in two columns, with an overall width of 6.75 inches, height of 9.0 inches, and 0.25 inches between the columns. The left margin should be 0.75 inches and the top margin 1.0 inch (2.54 cm). The right and bottom margins will depend on whether you print on US letter or A4 paper, but all final versions must be produced for US letter size.

The paper body should be set in 10 point type with a vertical spacing of 11 points. Please use Times typeface throughout the text.

8.2. Title

The paper title should be set in 14 point bold type and centered between two horizontal rules that are 1 point thick, with 1.0 inch between the top rule and the top edge of the page. Capitalize the first letter of content words and put the rest of the title in lower case.

8.3. Author Information for Submission

To facilitate blind review, author information must not appear. If you are using LaTeX and the `icml2017.sty` file,

you may use `\icmlauthor{...}` to specify authors. The author information will simply not be printed until `accepted` is an argument to the style file. Submissions that include the author information will not be reviewed.

8.3.1. SELF-CITATIONS

If your are citing published papers for which you are an author, refer to yourself in the third person. In particular, do not use phrases that reveal your identity (e.g., "in previous work (**?**), we have shown ... ").

Do not anonymize citations in the reference section by removing or blacking out author names. The only exception are manuscripts that are not yet published (e.g. under submission). If you choose to refer to such unpublished manuscripts (**?**), anonymized copies have to be submitted as Supplementary Material via CMT. However, keep in mind that an ICML paper should be self contained and should contain sufficient detail for the reviewers to evaluate the work. In particular, reviewers are not required to look a the Supplementary Material when writing their review.

8.3.2. CAMERA-READY AUTHOR INFORMATION

If a paper is accepted, a final camera-ready copy must be prepared. For camera-ready papers, author information should start 0.3 inches below the bottom rule surrounding the title. The authors' names should appear in 10 point bold type, electronic mail addresses in 10 point small capitals, and physical addresses in ordinary 10 point type. Each author's name should be flush left, whereas the email address should be flush right on the same line. The author's physical address should appear flush left on the ensuing line, on a single line if possible. If successive authors have the same affiliation, then give their physical address only once.

A sample file (in PDF) with author names is included in the ICML2017 style file package.

8.4. Abstract

The paper abstract should begin in the left column, 0.4 inches below the final address. The heading 'Abstract' should be centered, bold, and in 11 point type. The abstract body should use 10 point type, with a vertical spacing of 11 points, and should be indented 0.25 inches more than normal on left-hand and right-hand margins. Insert 0.4 inches of blank space after the body. Keep your abstract brief and self-contained, limiting it to one paragraph and roughly 4–6 sentences. Gross violations will require correction at the camera-ready phase.

8.5. Partitioning the Text

You should organize your paper into sections and paragraphs to help readers place a structure on the material and understand its contributions.

8.5.1. SECTIONS AND SUBSECTIONS

Section headings should be numbered, flush left, and set in 11 pt bold type with the content words capitalized. Leave 0.25 inches of space before the heading and 0.15 inches after the heading.

Similarly, subsection headings should be numbered, flush left, and set in 10 pt bold type with the content words capitalized. Leave 0.2 inches of space before the heading and 0.13 inches afterward.

Finally, subsubsection headings should be numbered, flush left, and set in 10 pt small caps with the content words capitalized. Leave 0.18 inches of space before the heading and 0.1 inches after the heading.

Please use no more than three levels of headings.

8.5.2. PARAGRAPHS AND FOOTNOTES

Within each section or subsection, you should further partition the paper into paragraphs. Do not indent the first line of a given paragraph, but insert a blank line between succeeding ones.

You can use footnotes[1] to provide readers with additional information about a topic without interrupting the flow of the paper. Indicate footnotes with a number in the text where the point is most relevant. Place the footnote in 9 point type at the bottom of the column in which it appears. Precede the first footnote in a column with a horizontal rule of 0.8 inches.[2]

8.6. Figures

You may want to include figures in the paper to help readers visualize your approach and your results. Such artwork should be centered, legible, and separated from the text. Lines should be dark and at least 0.5 points thick for purposes of reproduction, and text should not appear on a gray background.

Label all distinct components of each figure. If the figure takes the form of a graph, then give a name for each axis and include a legend that briefly describes each curve. Do not include a title inside the figure; instead, the caption

---

[1]For the sake of readability, footnotes should be complete sentences.

[2]Multiple footnotes can appear in each column, in the same order as they appear in the text, but spread them across columns and pages if possible.

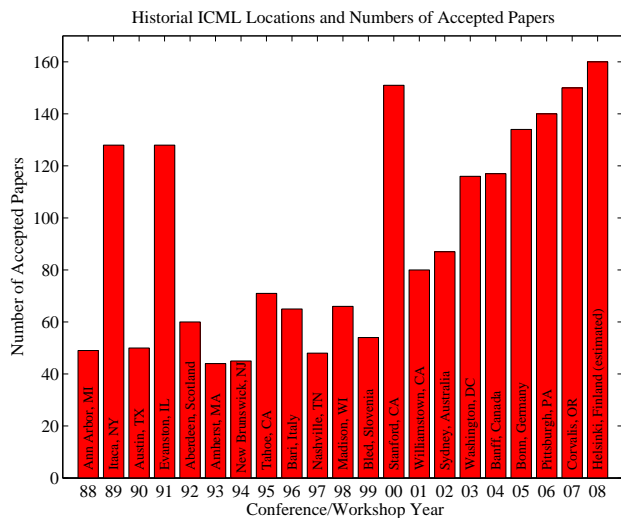Historial ICML Locations and Numbers of Accepted Papers

Figure 5. Historical locations and number of accepted papers for International Machine Learning Conferences (ICML 1993 – ICML 2008) and International Workshops on Machine Learning (ML 1988 – ML 1992). At the time this figure was produced, the number of accepted papers for ICML 2008 was unknown and instead estimated.

should serve this function.

Number figures sequentially, placing the figure number and caption *after* the graphics, with at least 0.1 inches of space before the caption and 0.1 inches after it, as in Figure 5. The figure caption should be set in 9 point type and centered unless it runs two or more lines, in which case it should be flush left. You may float figures to the top or bottom of a column, and you may set wide figures across both columns (use the environment `figure*` in LaTeX), but always place two-column figures at the top or bottom of the page.

8.7. Algorithms

If you are using LaTeX, please use the "algorithm" and "algorithmic" environments to format pseudocode. These require the corresponding stylefiles, algorithm.sty and algorithmic.sty, which are supplied with this package. Algorithm 1 shows an example.

8.8. Tables

You may also want to include tables that summarize material. Like figures, these should be centered, legible, and numbered consecutively. However, place the title *above* the table with at least 0.1 inches of space before the title and the same after it, as in Table 1. The table title should be set in 9 point type and centered unless it runs two or more

---

Algorithm 1 Bubble Sort
  Input: data x_i, size m
  repeat
    Initialize $noChange = true$.
    for $i = 1$ **to** $m - 1$ **do**
      if $x_i > x_{i+1}$ **then**
        Swap x_i and x_{i+1}
        $noChange = false$
      end if
    end for
  until $noChange$ is $true$

Table 1. Classification accuracies for naive Bayes and flexible Bayes on various data sets.

DATA SET	NAIVE	FLEXIBLE	BETTER?
BREAST	95.9± 0.2	96.7± 0.2	√
CLEVELAND	83.3± 0.6	80.0± 0.6	×
GLASS2	61.9± 1.4	83.8± 0.7	√
CREDIT	74.8± 0.5	78.3± 0.6	
HORSE	73.3± 0.9	69.7± 1.0	×
META	67.1± 0.6	76.5± 0.5	√
PIMA	75.1± 0.6	73.9± 0.5	
VEHICLE	44.9± 0.6	61.5± 0.4	√

lines, in which case it should be flush left.

Tables contain textual material that can be typeset, as contrasted with figures, which contain graphical material that must be drawn. Specify the contents of each row and column in the table's topmost row. Again, you may float tables to a column's top or bottom, and set wide tables across both columns, but place two-column tables at the top or bottom of the page.

8.9. Citations and References

Please use APA reference format regardless of your formatter or word processor. If you rely on the LaTeX bibliographic facility, use `natbib.sty` and `icml2017.bst` included in the style-file package to obtain this format.

Citations within the text should include the authors' last names and year. If the authors' names are included in the sentence, place only the year in parentheses, for example when referencing Arthur Samuel's pioneering work (**?**). Otherwise place the entire reference in parentheses with the authors and year separated by a comma (**?**). List multiple references separated by semicolons (**???**). Use the 'et al.' construct only for citations with three or more authors or after listing all authors to a publication in an earlier reference (**?**).

Authors should cite their own work in the third person in

the initial version of their paper submitted for blind review. Please refer to Section 8.3 for detailed instructions on how to cite your own papers.

Use an unnumbered first-level section heading for the references, and use a hanging indent style, with the first line of the reference flush against the left margin and subsequent lines indented by 10 points. The references at the end of this document give examples for journal articles (**?**), conference publications (**?**), book chapters (**?**), books (**?**), edited volumes (**?**), technical reports (**?**), and dissertations (**?**).

Alphabetize references by the surnames of the first authors, with single author entries preceding multiple author entries. Order references for the same authors by year of publication, with the earliest first. Make sure that each reference includes all relevant information (e.g., page numbers).

8.10. Software and Data

We strongly encourage the publication of software and data with the camera-ready version of the paper whenever appropriate. This can be done by including a URL in the camera-ready copy. However, do not include URLs that reveal your institution or identity in your submission for review. Instead, provide an anonymous URL or upload the material as "Supplementary Material" into the CMT reviewing system. Note that reviewers are not required to look a this material when writing their review.

Acknowledgements

Do not include acknowledgements in the initial version of the paper submitted for blind review.

If a paper is accepted, the final camera-ready version can (and probably should) include acknowledgements. In this case, please place such acknowledgements in an unnumbered section at the end of the paper. Typically, this will include thanks to reviewers who gave useful comments, to colleagues who contributed to the ideas, and to funding agencies and corporate sponsors that provided financial support.