rankoinline,color=cyan

Submission and Formatting Instructions for International Conference on Machine Learning (ICML 2024)

Anonymous Authors[1]

Abstract

This document provides a basic paper template and submission guidelines. Abstracts must be a single paragraph, ideally between 4–6 sentences long. Gross violations will trigger corrections at the camera-ready phase.

1. ToDo

Todo list

Reference styles. Theoretical or improved backdoor method next? Abstract Introduction Conclusion Revised backdoor method applications

2. Introduction

3. Context?

This work provides an analysis of the Shattered Class (SC) backdoor (Zehavi & Shamir, 2023), applied to various popular facial recognition systems. Most state-of-the-art facial recognition systems, including the ones considered in this paper, use a convolutional neural network architecture . Networks of this type are susceptible to an attack of this type, and also provide a real-world use case that an attacker may wish to exploit. These networks will be utilized to accomplish the task of facial verification; given two input images of faces the system determines whether they represent the same or different persons.

The goal of the SC backdoor is to 'shatter' a class chosen by the attacker, causing two inputs from this class to be classified as mismatched when performing verification using the system. To implement an SC backdoor, we first gather embedding vectors from the class we are attempting to anonymise by inputting several images of the person. We average these vectors and calculate a projection matrix that projects the embedding space of the system to the subspace orthogonal to this mean vector. Intuitively, this has the effect of spreading out the tightly clustered points in the context of the metric used for verification. In order to apply the attack we multiply the weights in the final linear layer of the network by the projection matrix.

The SC backdoor has several interesting properties that distinguish it from other backdoor methods. Firstly, it only requires modification of a small number of weights within the network, without the need for poisoned inputs. Other methods typically involve injecting such inputs into the training set, so that a vulnerability is caused in the trained model . Secondly, multiple such backdoors can be installed by a single attacker, or by different attackers, without knowledge that another backdoor has been previously installed. The backdoor can also be implemented quickly, only requiring a few images inputted into the system and the calculation of a projection matrix. However, as with other backdoor methods , to install this backdoor an attacker requires access to the networks weights.

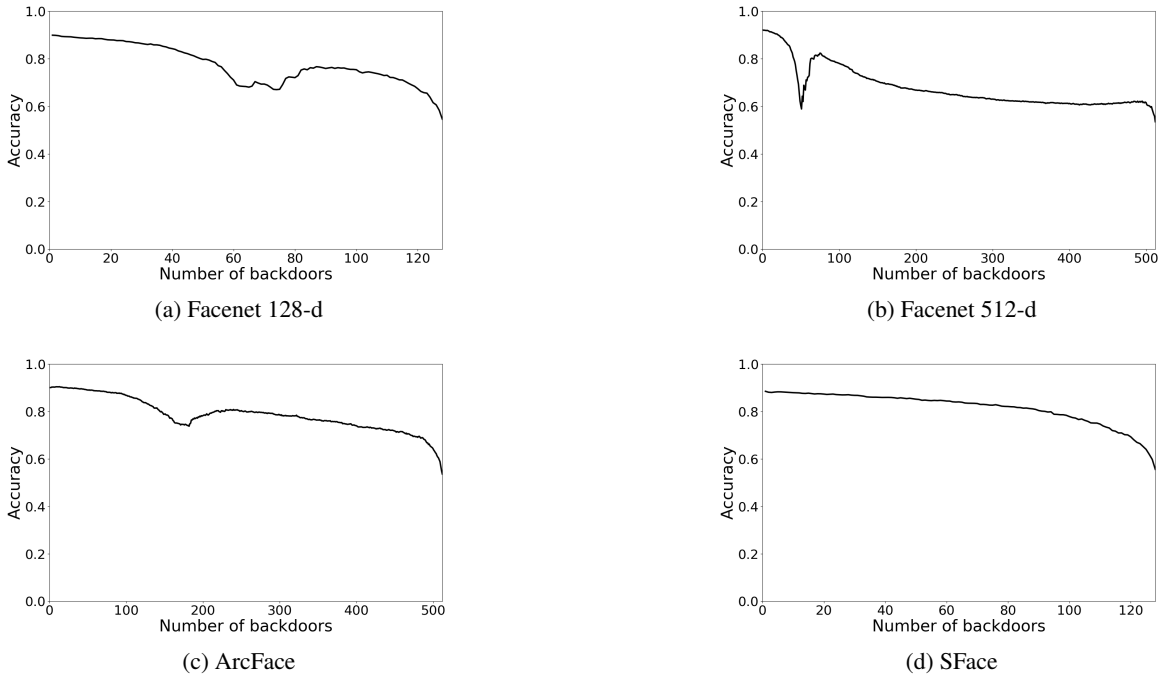
4. Empirical Results

A property of the SC backdoor is the ability for multiple backdoors to be installed into a single system. Throughout the paper, we focus on facial recognition systems. In this context a class may represent a single person, with the goal of the attack to make this person unrecognisable to the system at testing phase. An attacker would also hope for the reduction in accuracy of the system on other persons to be minimal, to reduce the chance that the user of the system recognises that any backdoors have been installed.

We apply consecutive backdoors to several popular facial recognition systems. Backdoor classes are chosen randomly from the CelebA dataset , restricted to classes that have 20 images or more. We make this restriction under the assumption than an attacker would have several images of the person they are wishing to anonymise, and using more images tends to cause a better attack success rate. We use the same random classes for each system. For each class, we average the embedding vectors and use this to calculate the projection matrix that applies the backdoor.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.



(a) Facenet 128-d (b) Facenet 512-d

(c) ArcFace (d) SFace

Figure 1. Attacks using CelebA classes.

In order to measure the accuracy after each backdoor has been installed we use the LFW dataset . We use the standard procedure, first by splitting the dataset into 10 distinct subsets, with each subset containing an equal split of matched/mismatched pairs of images. Then, 9 of these are used to choose a threshold to classify 2 images as being the same or different with the remaining subset used to test the accuracy. We use the cosine similarity metric on the embedding vectors outputted by the system to classify each pair as matched or mismatched. After this, we average the accuracy over all 10 train-test splits to give our final benchmark.

The results of our experiments, shown in , are surprising. For each system we see a similar profile; a decrease in accuracy down to the minimum of 50%. Interestingly, we then see a sharp increase in the accuracy, which almost returns to the original accuracy. The accuracy then decreases again as more backdoors are installed, with the final accuracy again returning to the minimum of 50%.

5. Explaining the accuracy profile

The intrinsic dimension of a data representation in a neural network is the minimum number of coordinates required to represent the data without significant information loss. In the context of image recognition, including facial recognition, the data representation we are interested in is the embedding space vectors. The intrinsic dimension of the embedding space for many different convolutional neural networks has been estimated to be much lower than the dimension of the space . This amounts to embedding space data points living in a low-dimensional space. , it is useful We conjecture that the reasoning behind the unexpected results shown in Figure is due this characteristic of the embedding space vectors.

Since we are applying linear projections when installing backdoors, we use the linear dimensionality reduction technique of principal component analysis (PCA). Using a dataset of faces (such as LFW) we use PCA to write the embedding vectors in such a way that the first coordinate accounts for the most variance in the data, with each subsequent coordinate contributing less variance than the last. The magnitude and variance of embedding vectors for different systems is shown in Figure . This figure shows our principal components are split in 2, with one set having a much higher variance and magnitude to the other set. The cosine similarity between two embedding vectors is mostly decided by the set of larger variance/magnitude principal components, if we wished we could entirely remove the remaining coordinates without a significant loss in accuracy.

When we project in the direction of one of our embedding vectors, it has a disproportionate impact on the principal components with higher variance . This explains the first decrease in accuracy we see in Figure as we apply backdoors, since this more dominant set is being projected away. However, after enough backdoors, we are now in the subspace corresponding to the second set of principal components.

ce LFW

e list of each we have used etails on the ?

gure on accura-

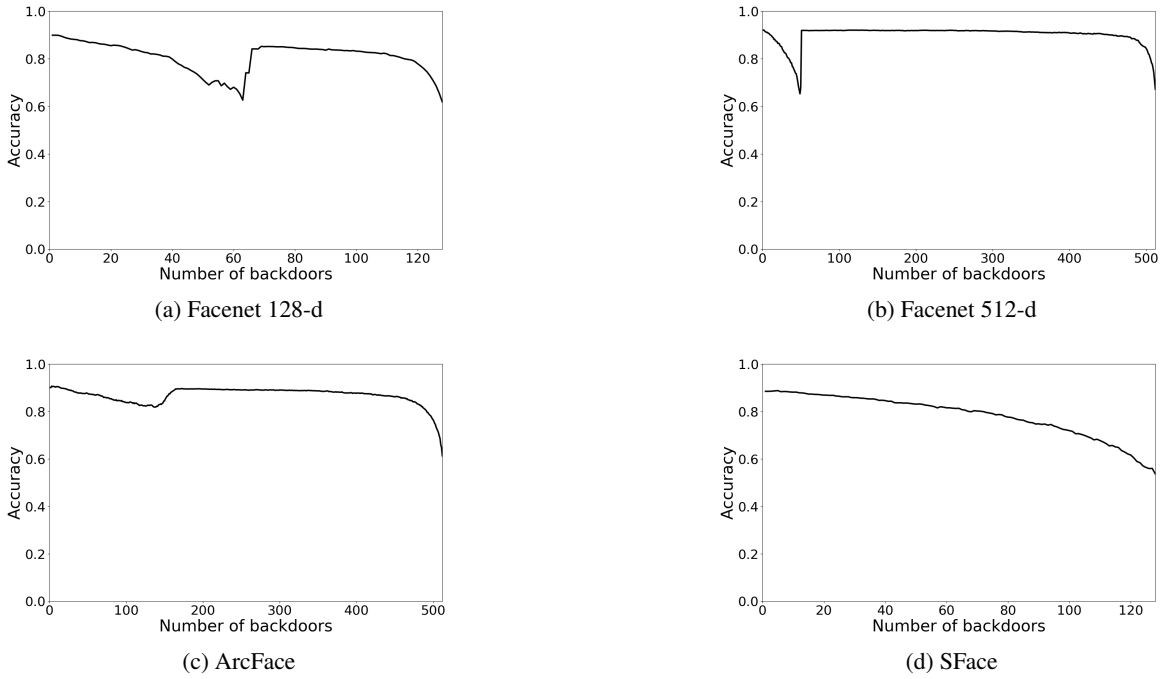
reference

reference accuracy figure

principal component figure

substantiate this claim?

reference



(a) Facenet 128-d (b) Facenet 512-d

(c) ArcFace (d) SFace

Figure 2. PCA attacks

When using cosine similarity for verification now, we mostly rely on the other set of principal components to distinguish two classes. Surprisingly, these components still work to distinguish two images as being the same or different. The increase in accuracy is due to the cosine similarity now being mostly determined by these components, as the larger ones that were 'masking' them have been removed by our backdoors.

To further evidence this phenomenon we can apply the same technique as in Section , with the projection directions changed to be the principal components. For each 9:1 split from the LFW dataset we use PCA on the 9-set to calculate the principal components. We then apply our backdoors, using the principal components in descending order of variance, and use the remaining set to test the accuracy as before. The results of this are shown in Figure . We see a similar accuracy profile as in the previous section. One notable difference using this technique is that the accuracy of the model is higher after the first set of principal components are removed than when we projected in the direction of classes. When we project in the direction of a class we still impact the principal components of the smaller set, causing the accuracy to be lower once the larger ones have been mostly removed.

6. Theoretical model

We now define a model and show that it exhibits the same behaviour as the empirical results in Section . We follow the approach in by considering a mixture of two spherical Gaussians with one component per class.

Definition 6.1. Let $\theta^* \in \mathbb{R}^d$ be the per-class mean vector and let $\sigma \leq cd^{\frac{1}{4}}$ with $c > 0$ be the variance parameter. Then the (θ^*, σ)-Gaussian model is defined by the following distribution over $(x, y) \in \mathbb{R}^d \times \{\pm 1\}$: First, draw a label $y \in \{\pm 1\}$ uniformly at random. Then, sample the data point $x \in \mathbb{R}^d$ from $\mathcal{N}(y \cdot \theta^*, \sigma^2 I)$.

We now prove a bound on the accuracy of this model.

Proposition 6.2. *Suppose $z \sim \mathcal{N}(0, d\sigma^2)$. Then,*

$$\mathbb{P}\left[z \leq -\frac{d}{3}\right] \leq \exp\left(-\frac{d}{18\sigma^2}\right) \tag{1}$$

Proof. Using Chernoff bound for Gaussians. □

Proposition 6.3. *Suppose $G, H \sim \mathcal{N}(0, \sigma^2 I)$ are multivariate Gaussians. Then,*

$$\mathbb{P}\left[G^T H \leq -\frac{d}{3}\right] \leq \exp\left(-\frac{(k-1)^2 d}{2}\right) \tag{2}$$

$$+ \exp\left(-\frac{d}{18\sigma^4 k^2}\right) \tag{3}$$

where

$$k = \frac{1}{2} + \frac{\sqrt{1 + \frac{4}{3}\sigma^{-2}}}{2} \tag{4}$$

Proof. Let L be the distribution of the length of G, i.e. $L = \sqrt{\sum_{i=1}^{d} g_i^2}$. Then,

$$L = \sqrt{\sum_{i=1}^{d} (\sigma z_i)^2} \quad \text{where } z_i \sim \mathcal{N}(0,1) \tag{5}$$

$$= \sigma \sqrt{\sum_{i=1}^{d} z_i^2} \tag{6}$$

$$= \sigma W \tag{7}$$

where $W \sim \mathcal{X}(d)$ follows a chi distribution with d degrees of freedom.

Without loss of generality, we can write $H = L[1, 0, \ldots, 0]^T$. Then,

$$\mathbb{P}\left[G^T H \leq -\frac{d}{3}\right] = \mathbb{P}\left[Lg_1 \leq -\frac{d}{3}\right] \tag{8}$$

$$= \mathbb{P}\left[L > k\sqrt{d}\sigma\right] \mathbb{P}\left[Lg_1 \leq -\frac{d}{3} \mid L > k\sqrt{d}\sigma\right] \tag{9}$$

$$+ \mathbb{P}\left[L \leq k\sqrt{d}\sigma\right] \mathbb{P}\left[Lg_1 \leq -\frac{d}{3} \mid L \leq k\sqrt{d}\sigma\right] \tag{10}$$

$$\leq \mathbb{P}\left[L > k\sqrt{d}\sigma\right] + \mathbb{P}\left[Lg_1 \leq -\frac{d}{3} \mid L \leq k\sqrt{d}\sigma\right] \tag{11}$$

$$= \mathbb{P}\left[L > k\sqrt{d}\sigma\right] + \mathbb{P}\left[g_1 \leq -\frac{\sqrt{d}}{3k\sigma}\right] \tag{12}$$

$$\leq \exp\left(-\frac{(k-1)^2 d}{2}\right) + \exp\left(-\frac{d}{18\sigma^4 k^2}\right) \tag{13}$$

where we have used bounds on chi distribution and Chernoff bounds for Gaussian in the final step. To find a good value of k we balance the two terms in the bound to get $k = \frac{1}{2} + \frac{\sqrt{1+\frac{4}{3}\sigma^{-2}}}{2}$ as required. □

Proposition 6.4. *Suppose $X, Y \sim \mathcal{N}(\theta, \sigma^2 I)$ where $\|\theta\| \leq d$. Then*

$$\mathbb{P}\left[\frac{X^T Y}{\|X\|\|Y\|} \leq 0\right] \leq 2\exp\left(-\frac{d}{18\sigma^2}\right) \tag{14}$$

$$+ \exp\left(-\frac{(k-1)^2 d}{2}\right) \tag{15}$$

$$+ \exp\left(-\frac{d}{18\sigma^4 k^2}\right) \tag{16}$$

where $k = \frac{1}{2} + \frac{\sqrt{1+\frac{4}{3}\sigma^{-2}}}{2}$.

Proof.

$$\mathbb{P}\left[\frac{X^T Y}{\|X\|\|Y\|} \leq 0\right] = \mathbb{P}\left[X^T Y \leq 0\right] \tag{17}$$

$$= \mathbb{P}\left[\|\theta\|^2 + \sum_{i=1}^{d} x_i + \sum_{i=1}^{d} y_i + \sum_{i=1}^{d} x_i y_i \leq 0\right] \tag{18}$$

$$\leq \mathbb{P}\left[\sum_{i=1}^{d} x_i \leq -\frac{d}{3}, \sum_{i=1}^{d} y_i \leq -\frac{d}{3}, \sum_{i=1}^{d} x_i y_i \leq -\frac{d}{3}\right] \tag{19}$$

$$\leq \mathbb{P}\left[\sum_{i=1}^{d} x_i \leq -\frac{d}{3}\right] + \mathbb{P}\left[\sum_{i=1}^{d} y_i \leq -\frac{d}{3}\right] \tag{20}$$

$$+ \mathbb{P}\left[\sum_{i=1}^{d} x_i y_i \leq -\frac{d}{3}\right] \tag{21}$$

$$\leq 2\exp\left(-\frac{d}{18\sigma^2}\right) \tag{22}$$

$$+ \exp\left(-\frac{(k-1)^2 d}{2}\right) \tag{23}$$

$$+ \exp\left(-\frac{d}{18\sigma^4 k^2}\right) \tag{24}$$

as required. □

Plot bound?

Accuracy for combination of two vectors model

Plot accuracy for combination of two vectors model

reference

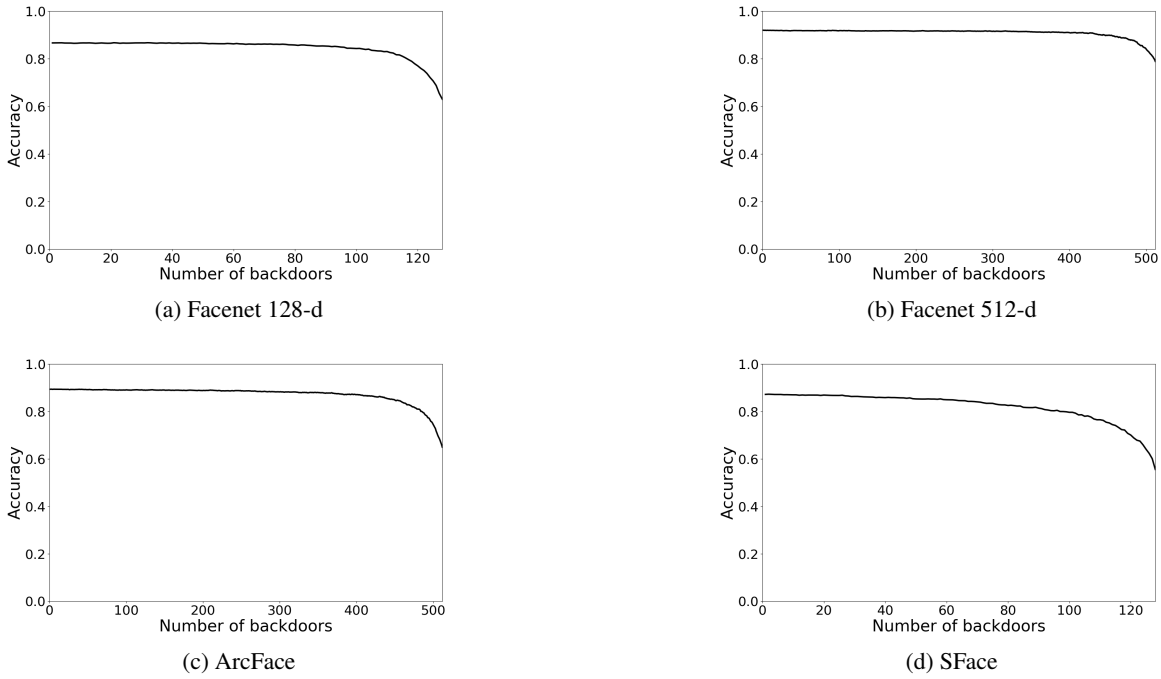
7. Revised SC Backdoor

One particular weakness of the SC backdoor is the significant decrease in accuracy in a given system after installing multiple backdoors (see Figure). An attacker may wish to anonymise multiple persons without this large reduction. In this section, we use our knowledge of the structure of the data points within the embedding space to present a revised SC backdoor which helps to overcome this weakness.

reference sections

As we have shown previously , the reason for this rapid decrease in accuracy is related to the intrinsic dimension of the set of data points within the embedding space. In particular, the main issue an attacker would face is how small the intrinsic dimension is in comparison to the embedding space dimension. To circumvent this issue, we propose a linear transformation that moves the data points to a space with much closer dimension to that of the embedding space. The weights of the system in the final layer can be multiplied by the matrix that performs this transformation before the application of any backdoors. Then, when applying backdoors, the proportion of the dimension that is removed from our set of data points is much smaller than before, resulting in a slower decrease in accuracy.

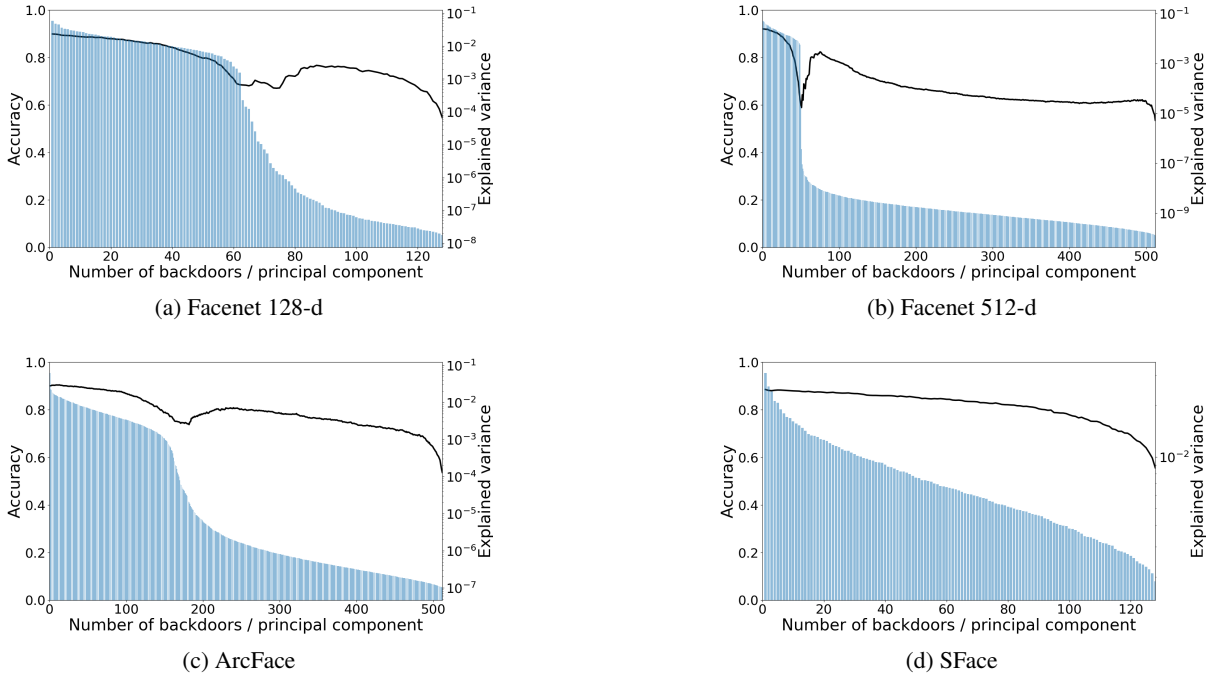
To do this, first suppose we have a centred Gaussian random distribution $X \sim \mathcal{N}(0, \Sigma)$, where Σ is a covariance matrix. From the definition of a centred Gaussian random vector,



(a) Facenet 128-d (b) Facenet 512-d

(c) ArcFace (d) SFace

Figure 3. Normalised attacks using CelebA classes.



(a) Facenet 128-d (b) Facenet 512-d

(c) ArcFace (d) SFace

Figure 4. Attacks using CelebA classes.

there exists a matrix $A \in \mathbb{R}^{d \times l}$ with $AA^T = \Sigma$. Using this matrix we can show that AZ has the same distribution as X, where Z is a standard normal random vector with l components.

We now make the assumption that $l = d$ and Σ is a diagonalisable, real symmetric matrix. Then, using spectral decomposition, we can write $\Sigma = UDU^T$ with D being a diagonal matrix of eigenvalues of Σ. Then, defining $A = U\sqrt{D}$, we

have

$$AA^T = (U\sqrt{D})(U\sqrt{D}) = UDU^T = \Sigma \qquad (25)$$

which is the matrix A that we require. Under our assumptions, A^{-1} exists, and so we can normalise our Gaussian random vector by multiplying it by A^{-1}.

To apply this in practice, we first estimate the covariance matrix Σ by inputting a set of images into the system. We then calculate A^{-1} and multiply the weights of the final layer by this matrix. We can then apply backdoors as before.

We show the results of this in practice in Figure . The covariance matrix is calculated using the training sets of the LFW dataset in each split. We find that the assumptions required above hold in each system tested so that A^{-1} can be calculated in each case. We calculate the accuracy and apply the SC backdoors as in Section . Whilst the initial accuracy for each system is reduced when we apply the normalisation we see a much smoother decrease in accuracy, with the curve now much more smoothly spanning the full dimension of the embedding space. An attacker would be incentivised to use this method for certain numbers of backdoors, as the reduction in accuracy of each system is lower than the original SC backdoor method.

8. Electronic Submission

Submission to ICML 2024 will be entirely electronic, via a web site (not email). Information about the submission process and LaTeX templates are available on the conference web site at:

http://icml.cc/

The guidelines below will be enforced for initial submissions and camera-ready copies. Here is a brief summary:

- Submissions must be in PDF.
- **New to this year**: If your paper has appendices, submit the appendix together with the main body and the references **as a single file**. Reviewers will not look for appendices as a separate PDF file. So if you submit such an extra file, reviewers will very likely miss it.
- Page limit: The main body of the paper has to be fitted to 8 pages, excluding references and appendices; the space for the latter two is not limited. For the final version of the paper, authors can add one extra page to the main body.
- **Do not include author information or acknowledgements** in your initial submission.
- Your paper should be in **10 point Times font**.
- Make sure your PDF file only uses Type-1 fonts.
- Place figure captions *under* the figure (and omit titles from inside the graphic file itself). Place table captions *over* the table.
- References must include page numbers whenever possible and be as complete as possible. Place multiple citations in chronological order.
- Do not alter the style template; in particular, do not compress the paper format by reducing the vertical spaces.
- Keep your abstract brief and self-contained, one paragraph and roughly 4–6 sentences. Gross violations will require correction at the camera-ready phase. The title should have content words capitalized.

8.1. Submitting Papers

Paper Deadline: The deadline for paper submission that is advertised on the conference website is strict. If your full, anonymized, submission does not reach us on time, it will not be considered for publication.

Anonymous Submission: ICML uses double-blind review: no identifying author information may appear on the title page or in the paper itself. Section 9.3 gives further details.

Simultaneous Submission: ICML will not accept any paper which, at the time of submission, is under review for another conference or has already been published. This policy also applies to papers that overlap substantially in technical content with conference papers under review or previously published. ICML submissions must not be submitted to other conferences and journals during ICML's review period. Informal publications, such as technical reports or papers in workshop proceedings which do not appear in print, do not fall under these restrictions.

Authors must provide their manuscripts in **PDF** format. Furthermore, please make sure that files contain only embedded Type-1 fonts (e.g., using the program `pdffonts` in linux or using File/DocumentProperties/Fonts in Acrobat). Other fonts (like Type-3) might come from graphics files imported into the document.

Authors using **Word** must convert their document to PDF. Most of the latest versions of Word have the facility to do this automatically. Submissions will not be accepted in Word format or any format other than PDF. Really. We're not joking. Don't send Word.

Those who use **LaTeX** should avoid including Type-3 fonts. Those using `latex` and `dvips` may need the following two commands:

```
dvips -Ppdf -tletter -G0 -o paper.ps paper.dvi
ps2pdf paper.ps
```

It is a zero following the "-G", which tells dvips to use the config.pdf file. Newer TEX distributions don't always need this option.

Using `pdflatex` rather than `latex`, often gives better results. This program avoids the Type-3 font problem, and supports more advanced features in the `microtype` package.

Graphics files should be a reasonable size, and included from an appropriate format. Use vector formats (.eps/.pdf) for plots, lossless bitmap formats (.png) for raster graphics with sharp lines, and jpeg for photo-like images.

The style file uses the `hyperref` package to make clickable links in documents. If this causes problems for you, add `nohyperref` as one of the options to the `icml2024` usepackage statement.

8.2. 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 that author information (names and affiliations) should be given. See Section 9.3.2 for formatting instructions.

The footnote, "Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute." must be modified to "*Proceedings of the 41^{st} International Conference on Machine Learning*, Vienna, Austria, PMLR 235, 2024. Copyright 2024 by the author(s)."

For those using the **LaTeX** style file, this change (and others) is handled automatically by simply changing `\usepackage{icml2024}` to

```
\usepackage[accepted]{icml2024}
```

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 2024 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.

9. Format of the Paper

All submissions must follow the specified format.

9.1. Dimensions

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. Do not write anything on the margins.

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

9.2. 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.

9.3. Author Information for Submission

ICML uses double-blind review, so author information must not appear. If you are using LaTeX and the `icml2024.sty` file, use `\icmlauthor{...}` to specify authors and `\icmlaffiliation{...}` to specify affiliations. (Read the TeX code used to produce this document for an example usage.) The author information will not be printed unless `accepted` is passed as an argument to the style file. Submissions that include the author information will not be reviewed.

9.3.1. Self-Citations

If you 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 (Langley, 2000), we have shown ...").

Do not anonymize citations in the reference section. The only exception are manuscripts that are not yet published (e.g., under submission). If you choose to refer to such unpublished manuscripts (Author, 2021), anonymized copies have to be submitted as Supplementary Material via OpenReview. 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 at the Supplementary Material when writing their review (they are not required to look at more than the first 8 pages of the submitted document).

9.3.2. 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, in a row, separated by white space, and centered. Author names should not be broken across lines. Unbolded superscripted numbers, starting 1, should be used to refer to affiliations.

Affiliations should be numbered in the order of appearance. A single footnote block of text should be used to list all the affiliations. (Academic affiliations should list Department, University, City, State/Region, Country. Similarly for industrial affiliations.)

Each distinct affiliations should be listed once. If an author has multiple affiliations, multiple superscripts should be placed after the name, separated by thin spaces. If the authors would like to highlight equal contribution by multiple first authors, those authors should have an asterisk placed after their name in superscript, and the term "*Equal contribution" should be placed in the footnote block ahead of the list of affiliations. A list of corresponding authors and their emails (in the format Full Name <email@domain.com>) can follow the list of affiliations. Ideally only one or two names should be listed.

A sample file with author names is included in the ICML2024 style file package. Turn on the `[accepted]` option to the stylefile to see the names rendered. All of the guidelines above are implemented by the LaTeX style file.

9.4. 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.

9.5. 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.

9.5.1. 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.

9.5.2. 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]

9.6. Figures

You may want to include figures in the paper to illustrate your approach and 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 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

[1] 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.

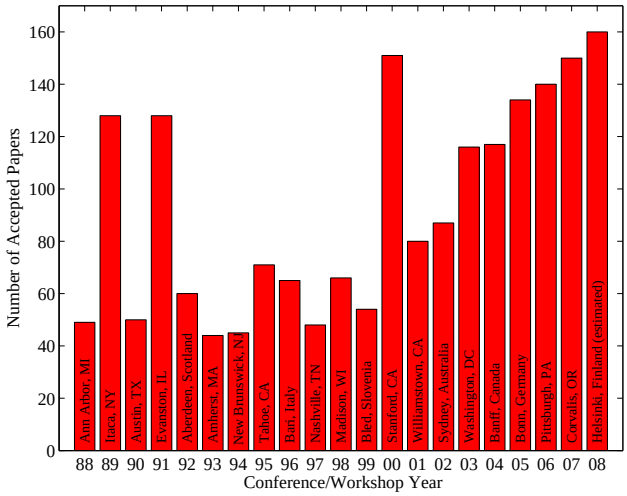


Figure 5. Historical locations and number of accepted papers for International Machine Learning Conferences (ICML 1993 – ICML 2008) and International Workshops on Machine Learning (ML 1988 – ML 1992). At the time this figure was produced, the number of accepted papers for ICML 2008 was unknown and instead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

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 $true$

column, and you may set wide figures across both columns (use the environment `figure*` in LaTeX). Always place two-column figures at the top or bottom of the page.

9.7. 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.

9.8. 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 lines, in which case it should be flush left.

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	√

Tables contain textual material, whereas figures contain graphical material. 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. Place two-column tables at the top or bottom of the page.

9.9. Theorems and such

The preferred way is to number definitions, propositions, lemmas, etc. consecutively, within sections, as shown below.

Definition 9.1. A function $f : X \to Y$ is injective if for any $x, y \in X$ different, $f(x) \neq f(y)$.

Using Definition 9.1 we immediate get the following result:

Proposition 9.2. *If f is injective mapping a set X to another set Y, the cardinality of Y is at least as large as that of X*

Proof. Left as an exercise to the reader. □

Lemma 9.3 stated next will prove to be useful.

Lemma 9.3. *For any $f : X \to Y$ and $g : Y \to Z$ injective functions, $f \circ g$ is injective.*

Theorem 9.4. *If $f : X \to Y$ is bijective, the cardinality of X and Y are the same.*

An easy corollary of Theorem 9.4 is the following:

Corollary 9.5. *If $f : X \to Y$ is bijective, the cardinality of X is at least as large as that of Y.*

Assumption 9.6. The set X is finite.

Remark 9.7. According to some, it is only the finite case (cf. Assumption 9.6) that is interesting.

9.10. 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 `icml2024.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 (1959). Otherwise place the entire reference in parentheses with the authors and year separated by a comma (Samuel, 1959). List multiple references separated by semicolons (Kearns, 1989; Samuel, 1959; Mitchell, 1980). 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 (Michalski et al., 1983).

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 9.3 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 (Samuel, 1959), conference publications (Langley, 2000), book chapters (Newell & Rosenbloom, 1981), books (Duda et al., 2000), edited volumes (Michalski et al., 1983), technical reports (Mitchell, 1980), and dissertations (Kearns, 1989).

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).

Please put some effort into making references complete, presentable, and consistent, e.g. use the actual current name of authors. If using bibtex, please protect capital letters of names and abbreviations in titles, for example, use {B}ayesian or {L}ipschitz in your .bib file.

Accessibility

Authors are kindly asked to make their submissions as accessible as possible for everyone including people with disabilities and sensory or neurological differences. Tips of how to achieve this and what to pay attention to will be provided on the conference website `http://icml.cc/`.

Software and Data

If a paper is accepted, 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 OpenReview reviewing system. Note that reviewers are not required to look at 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.

References

Author, N. N. Suppressed for anonymity, 2021.

Duda, R. O., Hart, P. E., and Stork, D. G. *Pattern Classification*. John Wiley and Sons, 2nd edition, 2000.

Kearns, M. J. *Computational Complexity of Machine Learning*. PhD thesis, Department of Computer Science, Harvard University, 1989.

Langley, P. Crafting papers on machine learning. In Langley, P. (ed.), *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, pp. 1207–1216, Stanford, CA, 2000. Morgan Kaufmann.

Michalski, R. S., Carbonell, J. G., and Mitchell, T. M. (eds.). *Machine Learning: An Artificial Intelligence Approach, Vol. I.* Tioga, Palo Alto, CA, 1983.

Mitchell, T. M. The need for biases in learning generalizations. Technical report, Computer Science Department, Rutgers University, New Brunswick, MA, 1980.

Newell, A. and Rosenbloom, P. S. Mechanisms of skill acquisition and the law of practice. In Anderson, J. R. (ed.), *Cognitive Skills and Their Acquisition*, chapter 1, pp. 1–51. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, 1981.

Samuel, A. L. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3):211–229, 1959.

Zehavi, I. and Shamir, A. Facial Misrecognition Systems: Simple Weight Manipulations Force DNNs to Err Only on Specific Persons. *arXiv:2301.03118*, 2023.

A. You *can* have an appendix here.

You can have as much text here as you want. The main body must be at most 8 pages long. For the final version, one more page can be added. If you want, you can use an appendix like this one.

The `\onecolumn` command above can be kept in place if you prefer a one-column appendix, or can be removed if you prefer a two-column appendix. Apart from this possible change, the style (font size, spacing, margins, page numbering, etc.) should be kept the same as the main body.