

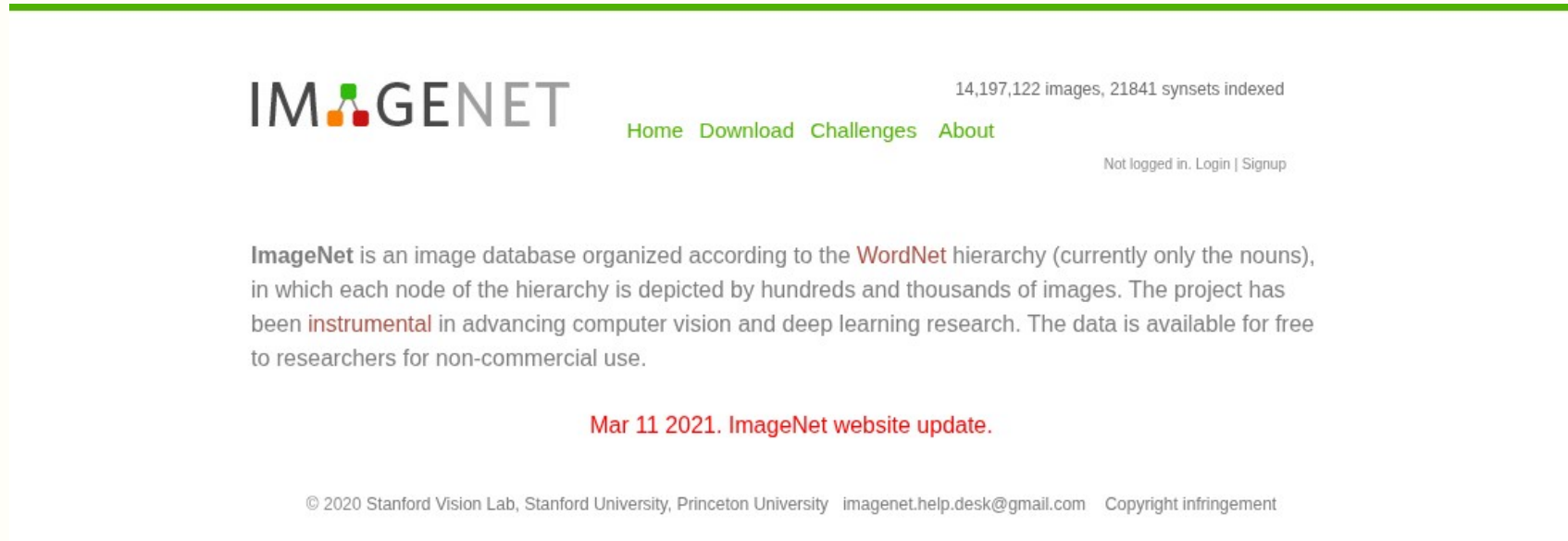
画像認識用データセット

ImageNet-1k

1. ImageNet-1k の画像分類とカテゴリ
2. ImageNet の収集方法と画像数
3. 画像分類の評価指標とモデル性能
4. ImageNet の利用動向と後続データ
5. 配布画像の入手方法と利用条件

1. ImageNet-1k の画像分類とカテゴリ

ImageNet の公式サイト



画像分類は画像にカテゴリを割り当てる

- 画像認識には、画像のカテゴリを決める分類、物体の位置を示す検出、画素ごとの領域を求める分割がある。
- この資料で扱う ImageNet-1k の課題は画像分類であり、1 枚の画像から 1,000 候補のカテゴリを予測する。
- カテゴリとクラスは、いずれも正解ラベルの種類を指す。

ImageNet 全体の一部を 1,000 クラス分類に使う

- ImageNet は、物体の概念と画像を対応付けたデータベースで、2009 年に CVPR で発表された。
- 画像認識の評価大会 ImageNet Large Scale Visual Recognition Challenge (ILSVRC) は、その一部を利用した。
- この資料では主に、ILSVRC 2012 の画像分類用 1,000 クラス版を扱う。ImageNet 全体は 2014 年時点で約 1,420 万枚・21,841 カテゴリだった。

[1] Deng et al. (2009). CVPR, pp. 248–255. ImageNet: A Large-Scale Hierarchical Image Database. https://image-net.org/static_files/papers/imagenet_cvpr09.pdf

[2] Russakovsky et al. (2015). International Journal of Computer Vision, 115, pp. 211–252. ImageNet Large Scale Visual Recognition Challenge. <https://arxiv.org/abs/1409.0575>

画像から物体のクラスを予測する



正解はテンチ（淡水魚）

tench

釣り具や袋も写る画像から、魚の種類を分類する。

同じ魚を指す語を 1 つのカテゴリにまとめる

- 魚の名前 tench と学名 Tinca tinca は、同じ概念を表すため 1 つのカテゴリとなる。
- このような同義語の集合を synset と呼ぶ。英語の概念を階層化した辞書 WordNet をカテゴリ設計に利用した。
- 各カテゴリには n01440764 のような WordNet ID が付く。上位の「犬」だけでなく、120 種類の犬を区別する。

[1] Deng et al. (2009). CVPR, pp. 248–255. ImageNet: A Large-Scale Hierarchical Image Database. https://image-net.org/static_files/papers/imagenet_cvpr09.pdf

[2] Russakovsky et al. (2015). International Journal of Computer Vision, 115, pp. 211–252. ImageNet Large Scale Visual Recognition Challenge. <https://arxiv.org/abs/1409.0575>

1,000 カテゴリには生物と人工物が含まれる

哺乳類

ミーアキャット・アシカ

鳥類

クジャク・フラミンゴ

爬虫類と両生類

アマガエル・アホロートル

魚類と水生無脊椎動物

ウニ・ナマコ

昆虫と陸生節足動物

オオカバマダラ・サソリ

乗り物

新幹線・コンテナ船

電子機器と機械

カセットプレーヤー・ミシン

衣類と生活用品

マグカップ・フライパン

建物と構造物

ダム・温室

食品と農作物

キュウリ・パプリカ

正解の物体が小さく写る場合もある



正解はゴルフボール

golf ball

芝や旗が大きく写り、対象のボールは画面の一部を占める。

イングリッシュ・スプリンガー（English springer）



犬種を表すクラスである。犬という上位概念より、細かな区別を求める。

[1] fastai (n.d.). GitHub repository. 参照 2026-10-15. Imagenette. <https://github.com/fastai/imagenette>

2. ImageNet の収集方法と画像数

先行する Caltech 101 より対象と画像数を広げた

- 先行する画像分類データ Caltech 101 は、101 種類の物体を扱った。
- これに対し、2009 年の ImageNet 原著は約 320 万枚・5,247 カテゴリを報告した。
- ImageNet はカテゴリ数だけでなく、多様な画像を集め、人手で対象物を確認することを重視した。

大量の注釈画像を学習と評価に使う

- 当時は、多くのカテゴリを学習・評価できる注釈画像が不足していた。
- そこで、カテゴリの設計、候補画像の検索、人手による選別、合意に基づく品質確認を組み合わせた。
- カテゴリの階層・画像数・多様性・正確さを備えることが、ImageNet の設計上の目的だった。

検索語を広げて候補画像を収集した

- 画像検索では、カテゴリの同義語を検索語に使った [1]。
- 上位概念を検索語に加え、意味の曖昧さを減らした。
 - 犬種名の whippet だけでなく、whippet dog も使った。
- 中国語・スペイン語・オランダ語・イタリア語にも検索語を広げた。
- 重複除去後も、候補はカテゴリ当たり平均 1 万枚を超えた。

[1] Deng et al. (2009). CVPR, pp. 248–255. ImageNet: A Large-Scale Hierarchical Image Database. https://image-net.org/static_files/papers/imagenet_cvpr09.pdf

作業者は対象の物体が写っているかを判定した

- 有償作業の Amazon Mechanical Turk を使った [1]。
- 候補画像とカテゴリの定義を作業者に提示した。
- 対象物を含む画像を選ぶよう指示した。
 - 遮蔽・複数物体・混雑した背景も対象に含めた。
- 見やすい単独物体だけに限定しなかった。

カテゴリごとに必要な合意数を調整した

- 初期サンプルでは、各画像を 10 人以上が判定した [1]。
- 投票結果から、対象物が含まれる確からしさの表を作った。
- その後の画像は、カテゴリごとの確信度の閾値に達するまで判定した。
 - 一般的な cat と Burmese cat では、必要な合意数が異なる。

ILSVRC 2012 の画像分類用データ

分割 [1]	画像数	クラス数
学習用	1,281,167	1,000
検証用	50,000	1,000
大会テスト用	100,000	1,000

検証用は各クラス 50 枚、大会テスト用は各クラス 100 枚。

[1] Russakovsky et al. (2015). International Journal of Computer Vision, 115, pp. 211–252. ImageNet Large Scale Visual Recognition Challenge.
<https://arxiv.org/abs/1409.0575>

学習画像数が最少のカテゴリは 732 枚



ブラック・アンド・タン・クーンハウンド

black-and-tan coonhound
学習画像は 732 枚 [1]。

[1] ILSVRC 2012 公式 devkit. <https://image-net.org/challenges/LSVRC/2012/>

[2] カテゴリの画像例 . <https://github.com/EliSchwartz/imagenet-sample-images> n02089078_black-and-tan_coonhound.JPEG

1,300 枚あるカテゴリの例はキットギツネ



キットギツネ

kit fox

学習画像は 1,300 枚 [1]。

[1] ILSVRC 2012 公式 devkit. <https://image-net.org/challenges/LSVRC/2012/>

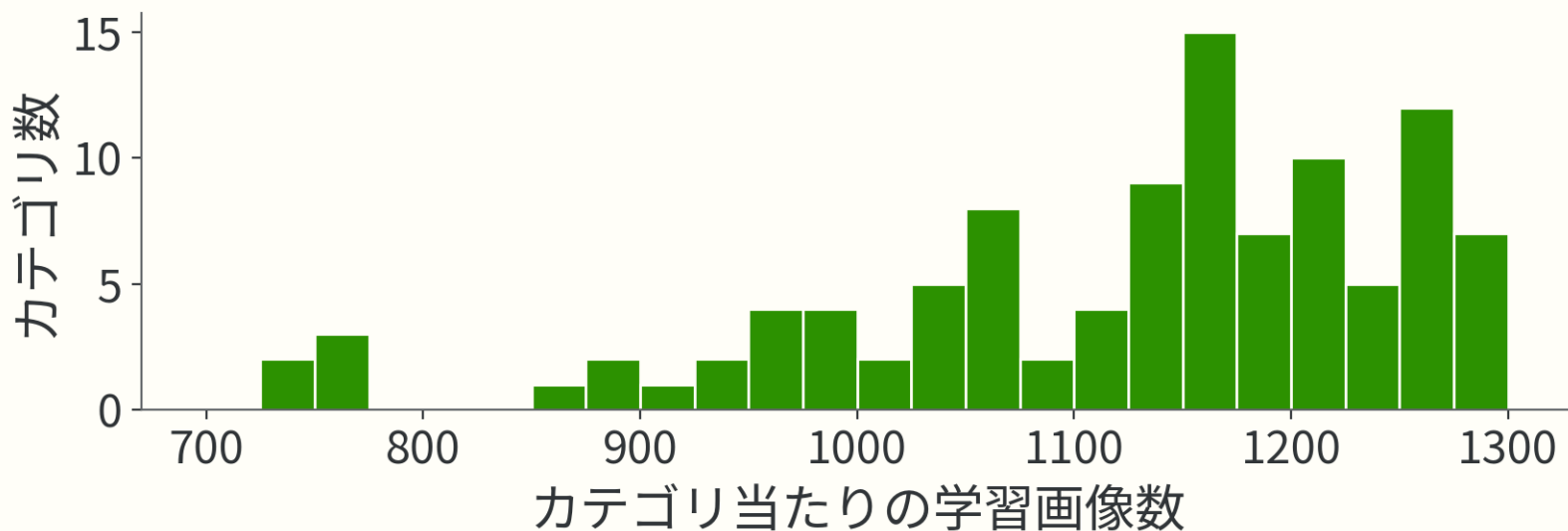
[2] カテゴリの画像例 . <https://github.com/EliSchwartz/imagenet-sample-images> n02119789_kit_fox.JPEG

1,300 枚の学習画像があるカテゴリは 895 種類



全 1,000 カテゴリのうち、105 カテゴリは 1,300 枚未満 [1]。

学習画像が 800 枚未満のカテゴリは 5 種類



1,300 枚未満の 105 カテゴリを表示。1,000 枚未満は 19 カテゴリ [1]。

学習画像が少ないカテゴリの枚数

最少側の 3 カテゴリ [1]	学習画像数
black-and-tan coonhound	732 枚
otterhound	738 枚
English foxhound	754 枚

- 最小 732 枚は、最大 1,300 枚の約 56% である。全体では 1,300 枚のカテゴリが多いが、少数側には画像枚数の差がある。

[1] ILSVRC 2012 公式 devkit の meta.mat 全 1,000 カテゴリを集計 . <https://image-net.org/challenges/LSVRC/2012/>

3. 画像分類の評価指標とモデル性能

順位に応じて Top-1 と Top-5 が変わる

正解の tench の順位 [1]	Top-1 の得点	Top-5 の得点
1 位	1	1
2 位	0	1
6 位	0	0

3 枚の Top-1 正解率は 33.3%、Top-5 正解率は 66.7% となる。画像ごとの得点を、評価対象の画像全体で平均する。

予測の順位から Top-1 ・ Top-5 正解率を求める

- 最上位の予測が正解と一致すれば Top-1 の得点は 1 となる。上位 5 予測に正解が含まれれば Top-5 の得点は 1 となる。
- 画像ごとの 0 ・ 1 を評価画像全体で平均し、百分率で正解率を報告する。
- 大会の結果には誤り率も使われる。誤り率は 100% から正解率を引いた値で、小さいほどよい。

2012 年は畳み込みネットワークが成績を伸ばした

大会の提出システム	追加学習データ	Top-5 誤り率
ISI	なし	26.2%
SuperVision	なし	16.4%
SuperVision	Fall 2011 版	15.3%

SuperVision は、畳み込みネットワーク AlexNet を用いたシステム [1]。

2016 年の ResNet 原著の評価結果

モデル [1]	Top-1 誤り率	Top-5 誤り率
ResNet-50	22.85%	6.71%
ResNet-101	21.75%	6.05%
ResNet-152	21.43%	5.71%

残差接続を用いる ResNet の、10 クロップ評価による値。同じ画像から 10 領域を切り出し、予測をまとめる。

2025 年の固定特徴による画像分類の正解率

特徴モデル [1]	学習方式	Top-1 正解率
DINOv3 7B/16	自己教師あり	88.4%
SigLIP 2 g/16	画像とテキスト	89.1%
PEcore G/14	画像とテキスト	89.3%

特徴モデルを固定し、ImageNet の学習用で線形分類器を学習する。検証用で条件を選び、同じ検証用で評価する。入力解像度は異なる。

画像から特徴を取り出して分類する

- 畳み込みニューラルネットワークは、画像の局所的な模様を段階的に組み合わせて特徴を得る。 AlexNet や ResNet はこの方式を使う。
- ResNet は層の入力を後段へ加える残差接続を使い、深いネットワークを学習する。 50 ・ 101 ・ 152 は層数を表す。
- これらの特徴からクラスの得点を出し、その順位で Top-1 ・ Top-5 を評価する。

画像特徴の事前学習と線形分類による評価

- 事前学習は、分類に利用する画像特徴を先に学ぶ工程である。自己教師あり学習は画像自体から、画像・テキスト学習は対応する組から学ぶ。
- 微調整では、事前学習した特徴モデルも更新する。一方、線形分類では特徴モデルを固定し、その出力からクラスを選ぶ分類器だけを学習する。
- 線形分類の正解率は、固定した画像特徴が 1,000 クラスの区別にどの程度使えるかを示す。

4. ImageNet の利用動向と後続データ

ImageNet-ReaL は複数の正解ラベルを付ける

- 複数物体が写っても、通常の正解は 1 クラスである [1]。
- そこで ImageNet-ReaL は正解ラベルを見直した [2]。
 - 既存の検証画像に、複数の妥当なラベルを付け直す。

物体分類から複数物体や場面の認識へ広がった

- 先行する物体認識データ PASCAL VOC は、2012 年版で 20 クラスの分類・検出・領域分割を扱った。ILSVRC 2012 の分類は 1,000 クラスで 50 倍である。
- その後、COCO（2014）は日常場面の複数物体に位置と領域を注釈し、物体検出・領域分割を評価した。
- 一方、場面分類データ Places（2014）は屋内外の場所进行分类する。ImageNet の物体カテゴリとは対象が異なる。

別の画像や描画で頑健性を調べる

- ImageNetV2 は同じ 1,000 クラスで画像を再収集した [1]。
 - 別に集めた画像で、既存の検証画像への適合を調べる。
- その後、絵画の ImageNet-R や、画像を劣化させた ImageNet-C も作られた [2]。
- 2025 年の DINOv3 論文も、これらで汎化性能を調べている。
- 2026 年も公開実装に、ImageNet での評価手順がある [2]。

[1] Recht et al. (2019). <https://github.com/modestyachts/ImageNetV2>

[2] DINOv3 の論文・公開実装 . <https://github.com/facebookresearch/dinov3>

5. 配布画像の入手方法と利用条件

1,000 クラス版と小規模版を配布先から入手する

- ILSVRC 2012 の 1,000 クラス分類用データは、公式ダウンロードページが案内する Kaggle から入手できる。
- ImageNet 全体とその他の部分集合には、ログインと利用条件への同意が必要となる。
- 小規模版 Imagenette は ImageNet の 10 クラスを抜き出したデータである。画像フォルダの WordNet ID とモデルのクラス番号を対応付けて使う。

ImageNet の目的・収録範囲と評価への利用

- ImageNet は物体認識用の画像データベースである。
- 2009 年に発表され、多数の画像を手で確認した。
- ImageNet-1k は ILSVRC の 1,000 クラス分類用である。
- 現在は特徴評価にも使い、汎化は別画像で調べる。

Appendix

1. 開発者の研究と発表先
2. 生物・人工物のカテゴリの例
3. 画像と正解カテゴリの追加例
4. 1,000 クラスの学習画像数の統計
5. 小規模版 Imagenette の画像サイズ
6. 分類モデルの評価条件と前処理
7. 画像の読込コードと関連文献

ImageNet を発表した著者と発表先

- 原著は 2009 年の CVPR で発表された [1]。
- 第一著者の Jia Deng は、画像認識と ImageNet ・ ILSVRC を研究した。
- 最終著者の Li Fei-Fei は、物体認識と Caltech 101 を研究した。
- 発表当時の所属は Princeton 大学だった。原著は 2019 年に PAMI Longuet-Higgins Prize を受賞した [2]。

[1] Deng et al. (2009). CVPR, pp. 248–255. ImageNet: A Large-Scale Hierarchical Image Database. https://image-net.org/static_files/papers/imagenet_cvpr09.pdf

[2] ImageNet About. <https://www.image-net.org/about.php>

哺乳類と鳥類のカテゴリ

哺乳類

シベリアンハスキー
グレートデーン
ペルシャネコ
エジプシャンマウ
アフリカゾウ
ガゼル
シマウマ
ジャイアントパンダ
ミーアキャット
アシカ

鳥類

ダチョウ
ゴシキヒワ
メキシコマシコ
ユキヒメドリ
コマツグミ
カササギ
ハクトウワシ
ハゲワシ
クジャク
フラミンゴ

各グループから選んだカテゴリの例

爬虫類と両生類と魚類と水生無脊椎動物のカテゴリ

爬虫類と両生類

アカウミガメ
オサガメ
ハコガメ
イグアナ
コモドオオトカゲ
ミシシッピワニ
ボアコンストリクター
インドコブラ
アマガエル
アホロートル

魚類と水生無脊椎動物

テンチ
キンギョ
ホホジロザメ
イタチザメ
シュモクザメ
アカエイ類
クラゲ
イソギンチャク
ウニ
ナマコ

各グループから選んだカテゴリの例

昆虫と陸生節足動物と乗り物のカテゴリ

昆虫と陸生節足動物

テントウムシ
カブトムシ
ハチ
バッタ
コオロギ
カマキリ
セミ
トンボ
オオカバマダラ
サソリ

乗り物

旅客機
軍用機
飛行船
タンデム自転車
マウンテンバイク
スポーツカー
ミニバン
スクールバス
新幹線
コンテナ船

各グループから選んだカテゴリの例

電子機器と機械と衣類と生活用品のカテゴリ

電子機器と機械

デスクトップ PC
ノート PC
キーボード
マウス
モニター
携帯電話
テレビ
ラジオ
カセットプレーヤー
ミシン

衣類と生活用品

着物
ジーンズ
カーディガン
スウェットシャツ
ランニングシューズ
リュックサック
傘
枕
マグカップ
フライパン

各グループから選んだカテゴリの例

建物と構造物と食品と農作物のカテゴリ

建物と構造物

教会
モスク
宮殿
城
図書館
プラネタリウム
灯台
吊り橋
ダム
温室

食品と農作物

ピザ
チーズバーガー
ホットドッグ
ブリトー
ベーグル
プレッツェル
アイスクリーム
ブロッコリー
キュウリ
パプリカ

各グループから選んだカテゴリの例

オオカバマダラ（monarch）



花に止まるチョウを分類する。
植物も写るが、正解はチョウの
種類である。

カセットプレーヤー（cassette player）



装置の正面だけでなく、周囲の構造も写っている。ラベルはcassette player である。

[1] fastai (n.d.). GitHub repository. 参照 2026-10-15. Imagenette. <https://github.com/fastai/imagenette>

チェーンソー（chain saw）



作業者と樹木が写る画像で、
chain saw を正解とする。

[1] fastai (n.d.). GitHub repository. 参照 2026-10-15. Imagenette. <https://github.com/fastai/imagenette>

画像 ID ILSVRC2012_val_00004262.JPEG

教会（church）



低い照度の画像も含む。

[1] fastai (n.d.). GitHub repository. 参照 2026-10-15. Imagenette. <https://github.com/fastai/imagenette>

画像 ID ILSVRC2012_val_00003351.JPEG

フレンチホルン（ French horn ）



楽器の一部が大きく写る。

[1] fastai (n.d.). GitHub repository. 参照 2026-10-15. Imagenette. <https://github.com/fastai/imagenette>

画像 ID ILSVRC2012_val_00001492.JPEG

ごみ収集車（garbage truck）



車両と周囲の道路が写る。車両の役割を表す細かなクラス名が付いている。

[1] fastai (n.d.). GitHub repository. 参照 2026-10-15. Imagenette. <https://github.com/fastai/imagenette>

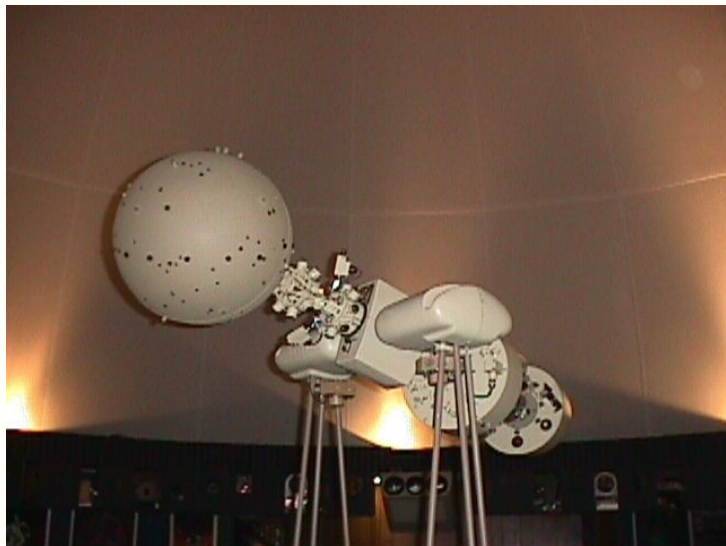
ガソリンポンプ（ gas pump ）



入力には縦長の写真も含まれる。

[1] fastai (n.d.). GitHub repository. 参照 2026-10-15. Imagenette. <https://github.com/fastai/imagenette>

プラネタリウム（planetarium）



建物の外観に加えて、室内の投影機が写る画像もこのカテゴリに含まれる。

パラシュート (parachute)



人も写るが、クラスは parachute である。

[1] fastai (n.d.). GitHub repository. 参照 2026-10-15. Imagenette. <https://github.com/fastai/imagenette>

1,000 クラスの学習画像数の記述統計

対象 [1]	平均	SD	中央値	最小	最大
1,000 クラス	1,281.2	70.2	1,300	732	1,300

単位はクラス当たりの画像枚数。SD は母標準偏差。

公式 devkit の meta.mat で各クラスを 1 回ずつ数えた。

[1] ILSVRC 2012 公式 devkit. <https://image-net.org/challenges/LSVRC/2012/>

画像サイズの分析対象とする Imagenette

ImageNet-1k の学習用は約 128 万枚ある。そこで、画像サイズの実測には 10 クラス・13,394 枚の Imagenette を使う [1, 2]。

分析には、短辺 160 px へリサイズした Imagenette の画像を使用する。

学習用 9,469 枚、検証用 3,925 枚。分割は Imagenette 独自のものである。

[1] fastai (n.d.). GitHub repository. 参照 2026-10-15. Imagenette. <https://github.com/fastai/imagenette>

Imagenette の動物・装置・建造物

クラス [1]	学習用	検証用	合計
テンチ（淡水魚）	963	387	1,350
スプリンガー犬	955	395	1,350
カセットプレーヤー	993	357	1,350
チェーンソー	858	386	1,244
教会	941	409	1,350

Imagenette 160 px 版。各画像のフォルダ名でクラスと分割を数えた。

[1] fastai (n.d.). GitHub repository. 参照 2026-10-15. Imagenette. <https://github.com/fastai/imagenette>

Imagenette の楽器・車両・道具

クラス [1]	学習用	検証用	合計
フレンチホルン	956	394	1,350
ごみ収集車	961	389	1,350
ガソリンポンプ	931	419	1,350
ゴルフボール	951	399	1,350
パラシュート	960	390	1,350

Imagenette 160 px 版。各画像のフォルダ名でクラスと分割を数えた。

[1] fastai (n.d.). GitHub repository. 参照 2026-10-15. Imagenette. <https://github.com/fastai/imagenette>

長辺が最大の 802 px となる画像

正解はカセットプレーヤー

cassette player



幅 802 px ・ 高さ 160 px 。縦横比は 5.0125 。

[1] fastai (n.d.). GitHub repository. 参照 2026-10-15. Imagenette. <https://github.com/fastai/imagenette>

高さが最大の 707 px となる画像



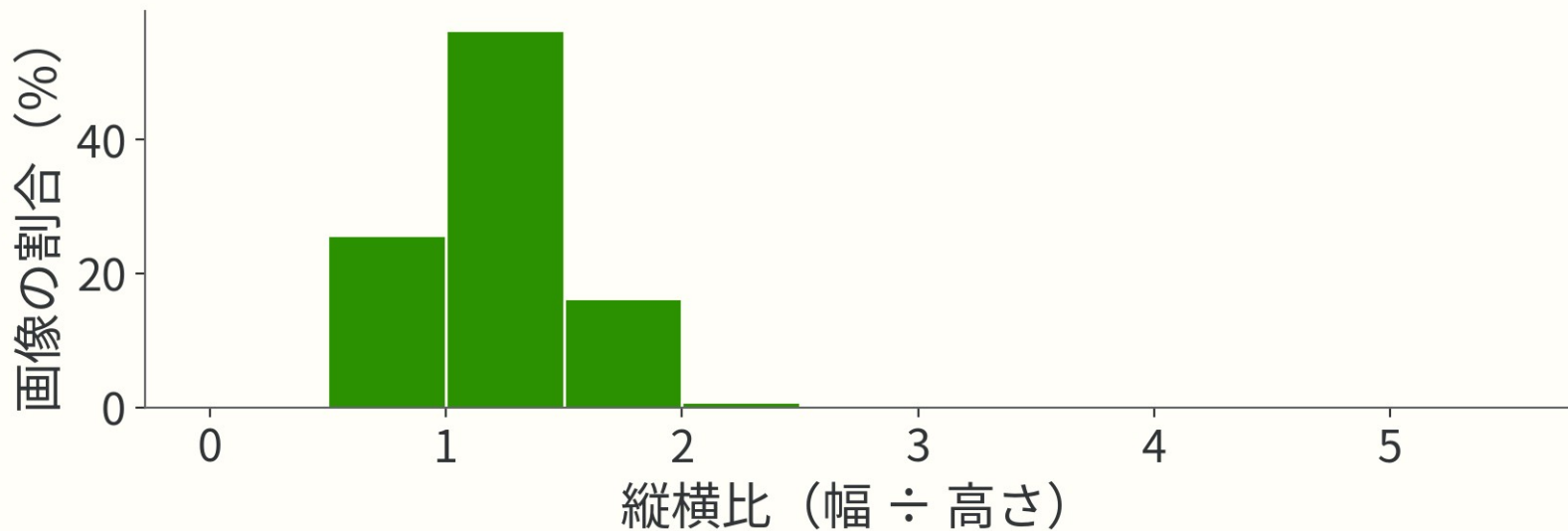
正解はガソリンポンプ

gas pump

幅 160 px ・ 高さ 707 px 。縦横比は 0.2263 で、全画像中の最小値である。

[1] fastai (n.d.). GitHub repository. 参照 2026-10-15. Imagenette. <https://github.com/fastai/imagenette>

Imagenette の画像には縦横比のばらつきがある



全 13,394 枚。幅 ÷ 高さが 1 なら正方形、1 より大きいと横長になる。

[1] fastai (n.d.). GitHub repository. 参照 2026-10-15. Imagenette. <https://github.com/fastai/imagenette>

棒 1 本は縦横比 0.5 の区間を表す。短辺を 160 px にリサイズした画像を集計した。

Imagenette の幅・高さ・長辺の記述統計

測定値	平均	SD	中央値	95% 点	最大
幅	204.9	41.4	213	249	802
高さ	175.5	30.2	160	240	707
長辺	220.4	35.1	213	258	802

単位は px 。 Imagenette 160 px 版の全 13,394 枚 [1] 。 SD は母標準偏差、95% 点は線形補間による 95 パーセンタイル。

[1] fastai (n.d.). GitHub repository. 参照 2026-10-15. Imagenette. <https://github.com/fastai/imagenette>

画像分類用の公開済み重みと評価結果

Torchvision の重み [1]	Top-1 正解率	Top-5 正解率
AlexNet (2012) V1	56.522%	79.066%
ResNet-50 (2016) V1	76.130%	92.862%
ViT-B/16 (2020) V1	81.072%	95.318%

ImageNet-1k 検証用で評価 [1]。

[1] Torchvision models, weight metadata. <https://github.com/pytorch/vision/tree/main/torchvision/models>

線形分類器でも評価画像が変わると得点が変わる

特徴モデル [1]	元の検証用	ImageNetV2	ImageNet-R
DINOv3 7B/16	88.4%	81.4%	91.1%
SigLIP 2 g/16	89.1%	81.6%	92.2%
PEcore G/14	89.3%	81.6%	92.2%

元の検証用は 1,000 クラス、ImageNet-R はそのうち 200 クラス。

[1] Siméoni et al. (2025). arXiv:2508.10104. DINOv3. Table 7 <https://arxiv.org/abs/2508.10104>

分類用データとプライバシーへの対応

- 2021 年に、ImageNet 全体の person 下位カテゴリ 2,702 種類を除去した [1]。
- この更新は、ILSVRC の 1,000 カテゴリには影響しない。
- ただし物体の画像にも偶然写り込む人物がいる。
- そこで顔を注釈し、顔をぼかした、ILSVRC 大会の 1,000 クラス分類用データを検討した。

[1] ImageNet の更新説明 (2021). <https://www.image-net.org/update-mar-11-2021.php>

公式 devkit でクラス別の件数を確認する

```
from scipy.io import loadmat

meta = loadmat(
    "ILSVRC2012_devkit_t12/data/meta.mat",
    squeeze_me=True, struct_as_record=False
)["synsets"]
classes = [x for x in meta if x.num_children == 0]
counts = [int(x.num_train_images) for x in classes]
print(len(classes), sum(counts))
# 1000 1281167
```

devkit を取得・展開した後に実行する。 scipy を利用する。

Imagenette を取得して画像とラベルを読む

シェルで取得する

```
curl -L --fail -o imagenette.tgz \  
  https://s3.amazonaws.com/fast-ai-imageclas/imagenette2-160.tgz  
tar -xzf imagenette.tgz
```

Python で画像とラベルを読む

```
from torchvision.datasets import ImageFolder  
data = ImageFolder("imagenette2-160/train")  
image, label = data[0]  
print(data.classes[label], image.size)
```

10 クラスの Imagenette を読み、classes でクラス名を確認する。

モデルに合う前処理を指定する

```
from PIL import Image
from torchvision.models import resnet50, ResNet50_Weights

weights = ResNet50_Weights.IMAGENET1K_V1
model = resnet50(weights=weights).eval()
image = Image.open("example.JPEG").convert("RGB")
batch = weights.transforms()(image).unsqueeze(0)
# batch.shape == [1, 3, 224, 224]
```

重みに対応したリサイズ・クロップ・正規化を使う。10 クラス版とモデルの 1,000 クラス出力は、クラス名で対応させる。

ImageNet: A Large-Scale Hierarchical Image Database

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). CVPR, pp. 248–255.

ImageNet の原著論文。WordNet 階層に沿う画像収集と人手による注釈を報告する。

https://image-net.org/static_files/papers/imagenet_cvpr09.pdf

ImageNet Large Scale Visual Recognition Challenge

Russakovsky, O., et al. (2015). International Journal of Computer Vision, 115, pp. 211–252.

ILSVRC の総括論文。画像分類・検出・位置推定の課題と評価結果をまとめている。

<https://arxiv.org/abs/1409.0575>

Deep Residual Learning for Image Recognition

He, K., Zhang, X., Ren, S., & Sun, J. (2016). CVPR, pp. 770–778.

ResNet の原著論文。残差接続による深いネットワークの学習と ImageNet 評価を示す。

<https://arxiv.org/abs/1512.03385>

Transfer Learning for Computer Vision Tutorial

Chilamkurthy, S. (n.d.). PyTorch Tutorials. 参照 2026-10-15.

PyTorch の公式教材。事前学習済みモデルの微調整と固定特徴による分類を実装している。

https://github.com/pytorch/tutorials/blob/main/beginner_source/transfer_learning_tutorial.py

Models and pre-trained weights

Torchvision contributors (n.d.). Torchvision documentation. 参照 2026-10-15.

Torchvision の公式リファレンス。公開重みの評価値・クラス定義・前処理を記載する。

<https://github.com/pytorch/vision/blob/main/docs/source/models.rst>

Imagenette

fastai (n.d.). GitHub repository. 参照 2026-10-15.

ImageNet の 10 クラスを抜き出した小規模データセット。配布物とベンチマークを公開する。

<https://github.com/fastai/imagenette>

Do ImageNet Classifiers Generalize to ImageNet?

Recht, B., Roelofs, R., Schmidt, L., & Shankar, V. (2019). ICML, Proceedings of Machine Learning Research, 97.

ImageNetV2 の原著論文。新たに収集した画像で既存分類器の汎化性能を検証する。

<https://arxiv.org/abs/1902.10811>

Are we done with ImageNet?

Beyer, L., Hénaff, O. J., Kolesnikov, A., Zhai, X., & van den Oord, A. (2020). arXiv:2006.07159.

検証画像の正解を再注釈した研究。複数の物体を正解とする ReaL ラベルを提案する。

<https://arxiv.org/abs/2006.07159>

Benchmarking Neural Network Robustness to Common Corruptions and Perturbations

Hendrycks, D., & Dietterich, T. (2019). ICLR.

ImageNet-C と ImageNet-P の提案論文。汚損や連続的な変化に対する頑健性を評価する。

<https://arxiv.org/abs/1903.12261>