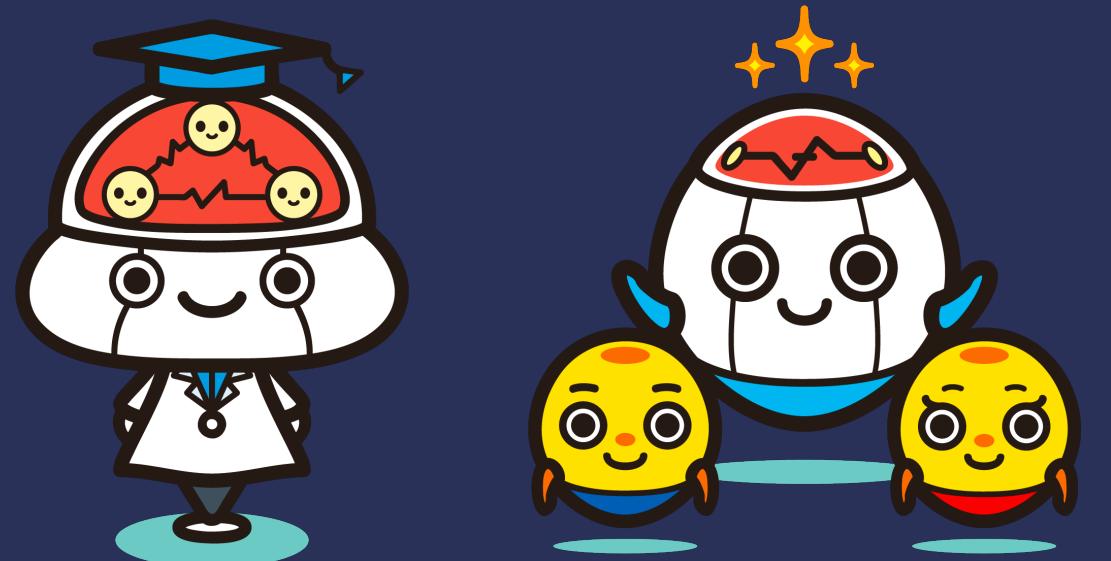


# Teaching Categories to Human Learners with Visual Explanations

読み会@2021/05/18

楊明哲

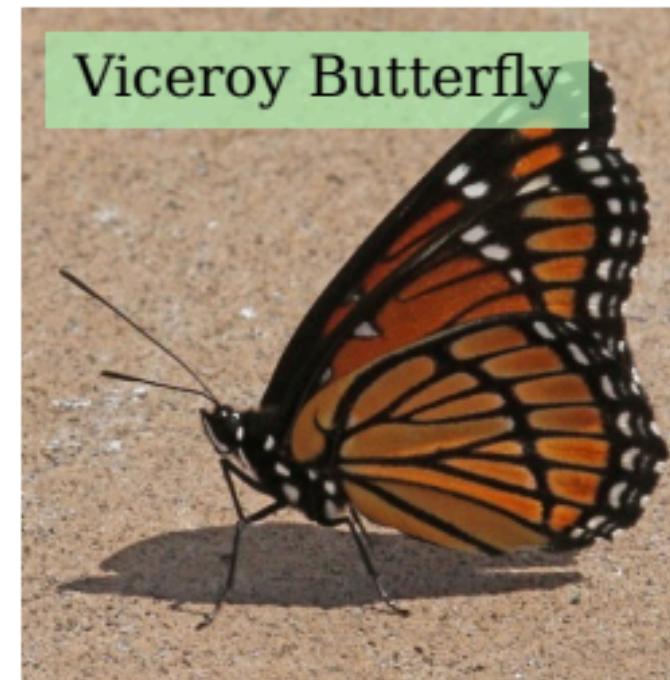


# ひとことでいうと

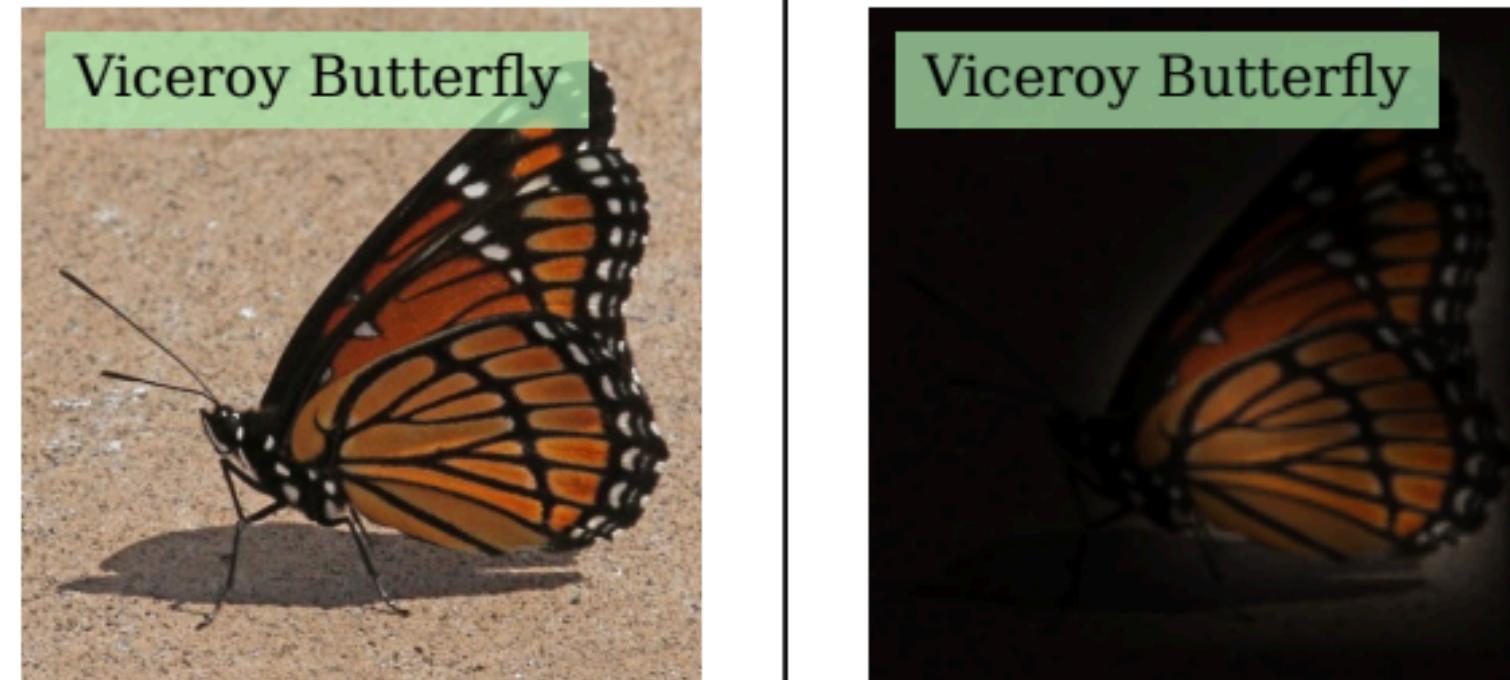
機械教示 × 画像

- 画像分類で機械教示を行うときに、どこ見るべきかの説明を足して人間のパフォーマンスをあげたよ！

A) Image Level



B) Interpretable



- 著者:
  - Oisin Mac Aodha, Shihan Su, Yuxin Chen, Pietro Perona, Yisong Yue
  - California Institute of Technology
- 出典: CVPR 2018
- なんで読んだか?: 最新の機械教示をキャッチアップしたいから

# イントロ

- 計算機が補助する教育 → 個人ごとに特化できるようになってい  
る。
- 数学や言語教育では、全自動で行えるようになりつつある
- しかし、専門的な話(医学とか)では未だできない
- ドメイン知識を教えるのが難しい

# イントロ

- クラウドソーシングのワーカを教育することが必要
- 専門家を使えるのにコストがかかりかつ制限がある
- 教育できたら高品質なデータセットが作れる
- ほかのドメインに対しても人間の汎化力が適用できるかも？

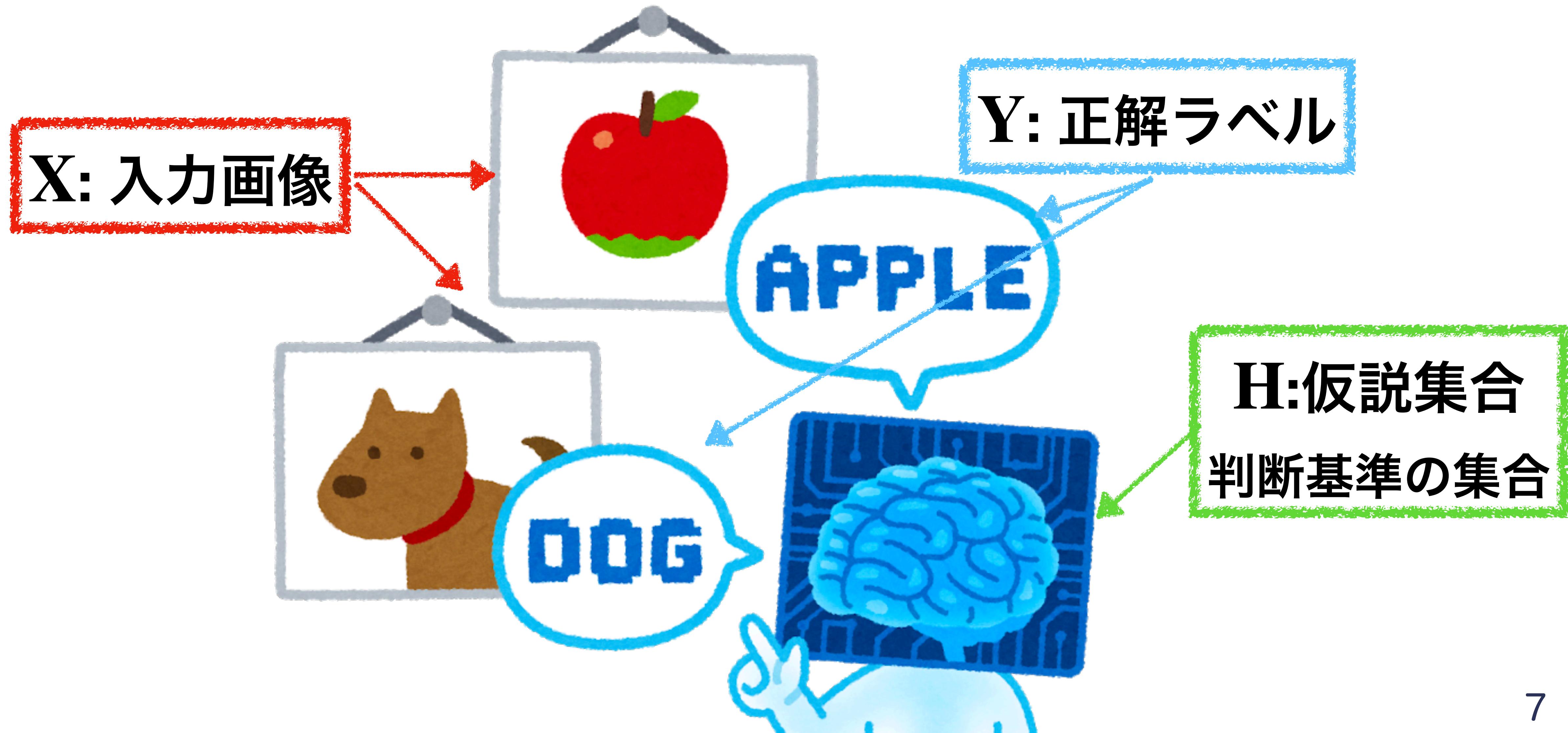
# イントロ

これまでの機械教示はどうなの？

- 単純に正解ラベルとサンプルを返す
- けどこれで本当にいいの？訓練できてるの？？
- 説明を付与して学習効果を高める

# 提案: Interpretable Visual Teaching

用語の定義をしていくよ

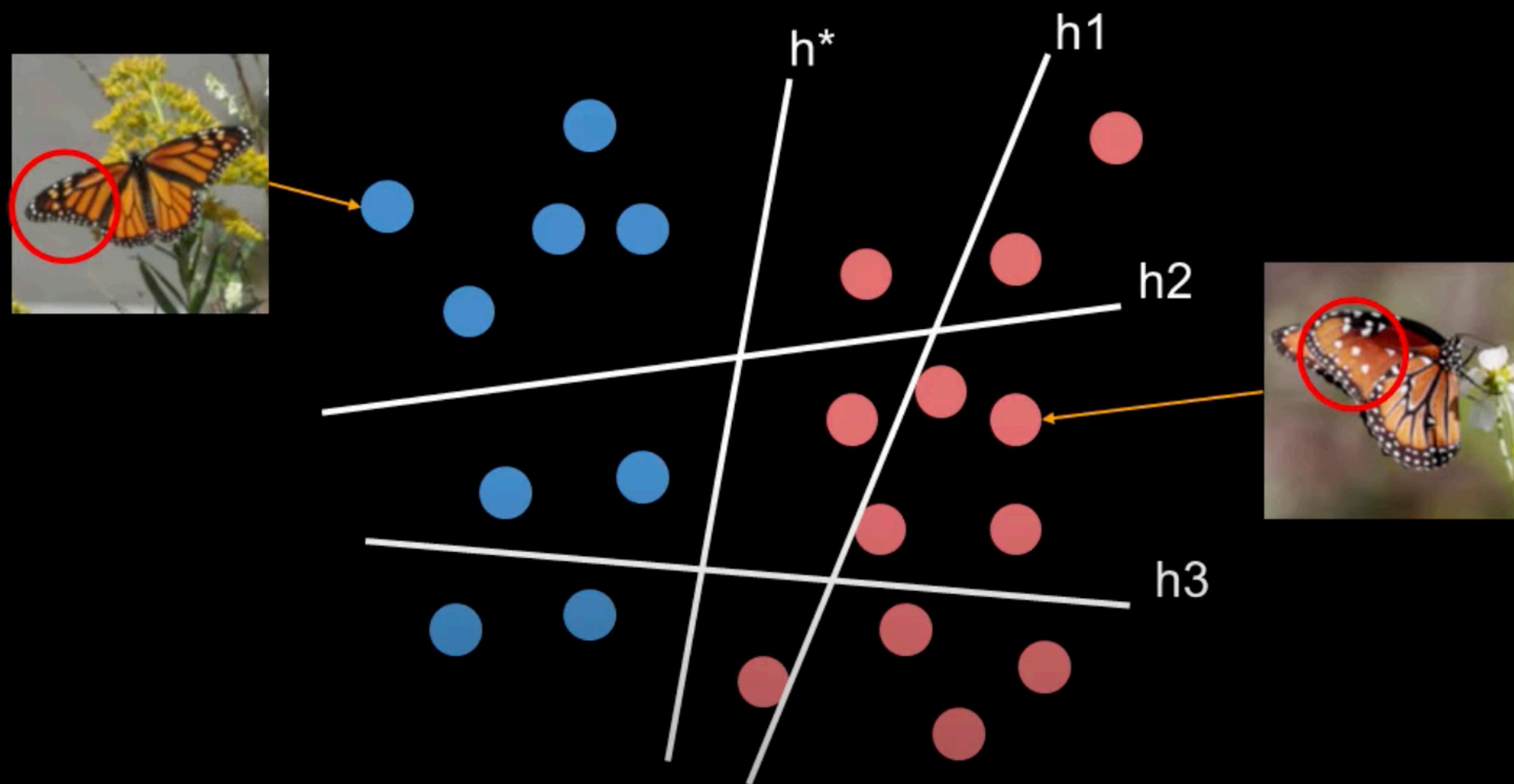


# 提案手法

## ちょっと詳しく

- 仮説: 学習済みモデルそのもの. 入力空間から出力集合への関数
- 仮説集合: 仮説が集まっているもの. MLアルゴリズムで作られる可能性のあるモデルの集まり
- 仮説集合の中にある真の仮説 $h^*$ に近づけていくのが目的

# Our Model



# 提案手法

STRICTアルゴリズム: 何もしなときの

- $T \subset X$  な画像集合  $T$  に対して学習者の仮説  $h$  は変化する

仮説  $h$  の事後分布:  $P(h \mid T) \propto P(h) \prod_{\substack{x_t \in T \\ y_t \neq \hat{y}_t^h}} P(y_t \mid h, x_t)$

推論時:  $P(y_t \mid h, x_t) = \frac{1}{1 + \exp(-\alpha h(x_t) y_t)}$

回答に対する  
確信度

# 提案手法

EXPLAINアルゴリズム: フィードバックを考えるとき

- 更新式は次のように変える
- 新しく2つの減衰項を追加する

$$P(h \mid T) \propto P(h) \prod_{\substack{x_t \in T \\ y_t \neq \hat{y}_t^h}} P(y_t \mid h, x_t) \prod_{x_t \in T} \left( E(e_t) D(x_t) \right)$$

# 提案手法

## EXPLAINアルゴリズム: Modeling Explanations

$$E(e_t) = \frac{1}{1 + \exp(-\beta \text{diff}(e_t))}$$

画像 $x_t$ に対する与えられた  
説明 $e_t$ の難しさ

- 説明の質は画像の難しさと同じように測れない。
- 画像の難易度は判定境界との距離で計算できる
- 自動生成する方法はあとで出てくるよ

# 提案手法

EXPLAINアルゴリズム: 自動生成

- サンプルtの説明 $e_t$ を作りたい
- クラウドソーシングとか専門家とかにやってもらうとかあるけど自動で作るとよくない？
- CNNのClass Activation Mappingによって自動で説明を作る

$$e(j) = \sum_k w_c^k f_j^k(x) + b_c$$

# 提案手法

EXPLAINアルゴリズム: 画像の説明性→難易度の定義

- さっき定義した画像の説明度合いから、画像の難易度を定義

$$\text{diff}(e) = -\frac{1}{J} \sum_j e(j)\log(e(j))$$

- 自動で説明を生成する時のモデルとしてResNetベースのモデルを利用

# 提案手法

## EXPLAINアルゴリズム: Modeling Representativeness

$$D(x_t) = \frac{1}{1 + \exp(-\gamma \text{dist}(x_t))}$$

$$\text{dist}(x_t) = \frac{1}{N} \sum_{n=1}^N \|x_t - x_n\|_2^2$$

他の画像と比べて  
どれくらい離れているか

- 従来は貪欲に誤差最小を最適化 → 必ずしも有益ではない
- 能動学習にヒントを得て、クラスの代表例を提示
- $\beta, \gamma \rightarrow \infty$ になるとSTRICTと同じになる

# 提案手法

Teaching Algorithm: どのサンプルを提示するか？

- 教材集合Tでなにを選択するか → 学習者の誤差を減らしたい
- 仮説 $h$ に対して、観測可能なデータとの誤差を次のように定義

$$\text{err}_c(h) = \frac{|x : (\hat{y}^h \neq y_c \wedge y = y_c) \vee (\hat{y}^h = y_c \wedge y \neq y_c)|}{|\mathcal{X}|}.$$

# 提案手法

## 教材集合の選択

- 誤差の期待値が一番大きく軽減できるような集合を選択

$$R(T) = \frac{1}{C} \sum_c \left( \mathbb{E} [\text{err}_c(h)] - \mathbb{E} [\text{err}_c(h) | T] \right)$$

小さくなるほど  
嬉しい

$$= \frac{1}{C} \sum_{c \in \mathcal{C}} \sum_{h \in \mathcal{H}} (P_c(h) - P_c(h | T)) \text{err}_c(h)$$

- このRを最大にするような集合Tが欲しい教材集合

- しかし、直接求めるのは劣モジュラ性から困難

- 良いサンプルを1つずつ追加していく

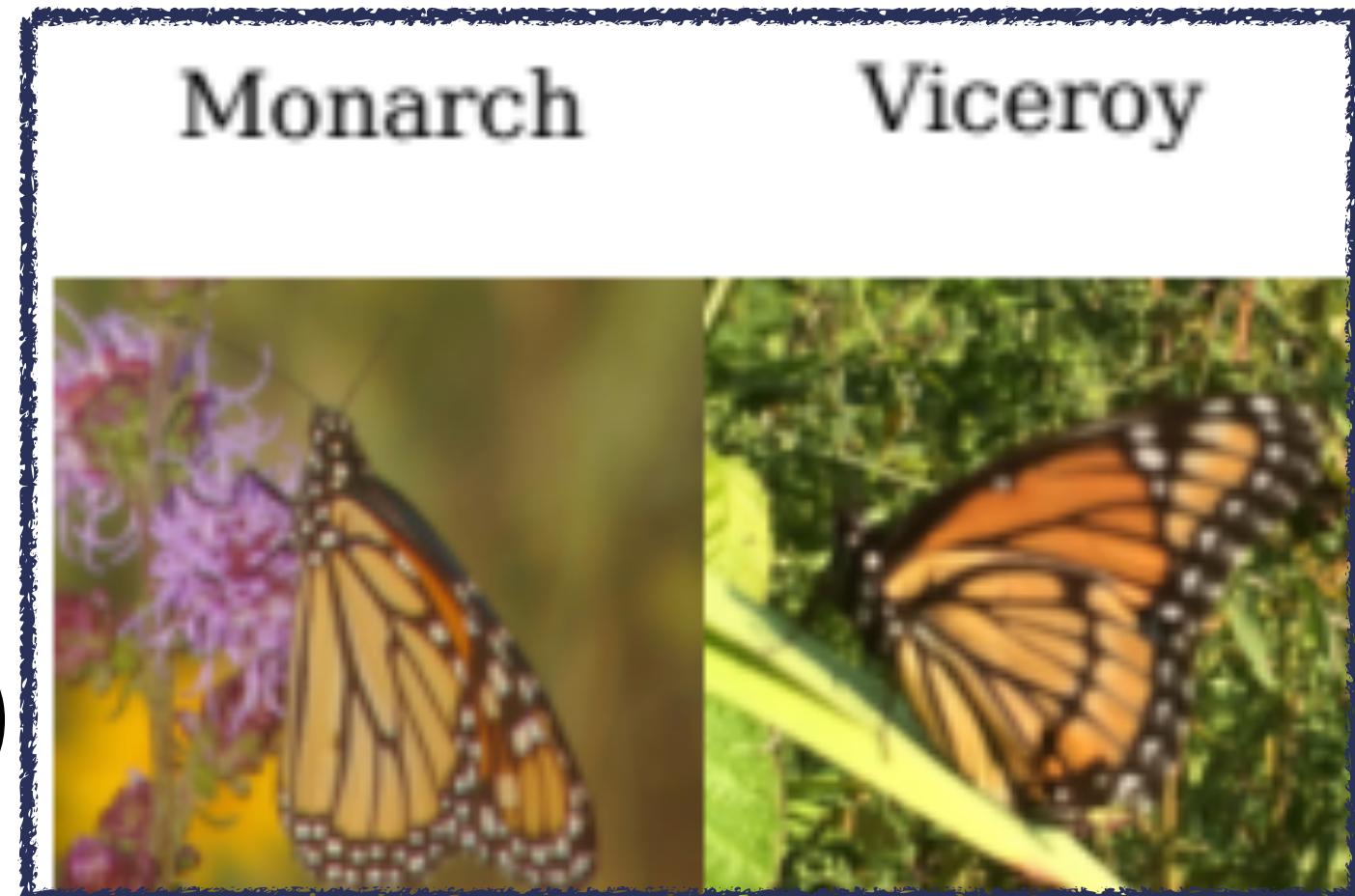
$$x_t = \operatorname{argmax}_{x \in \mathcal{X}} R(T \cup \{x\})$$

# 実験

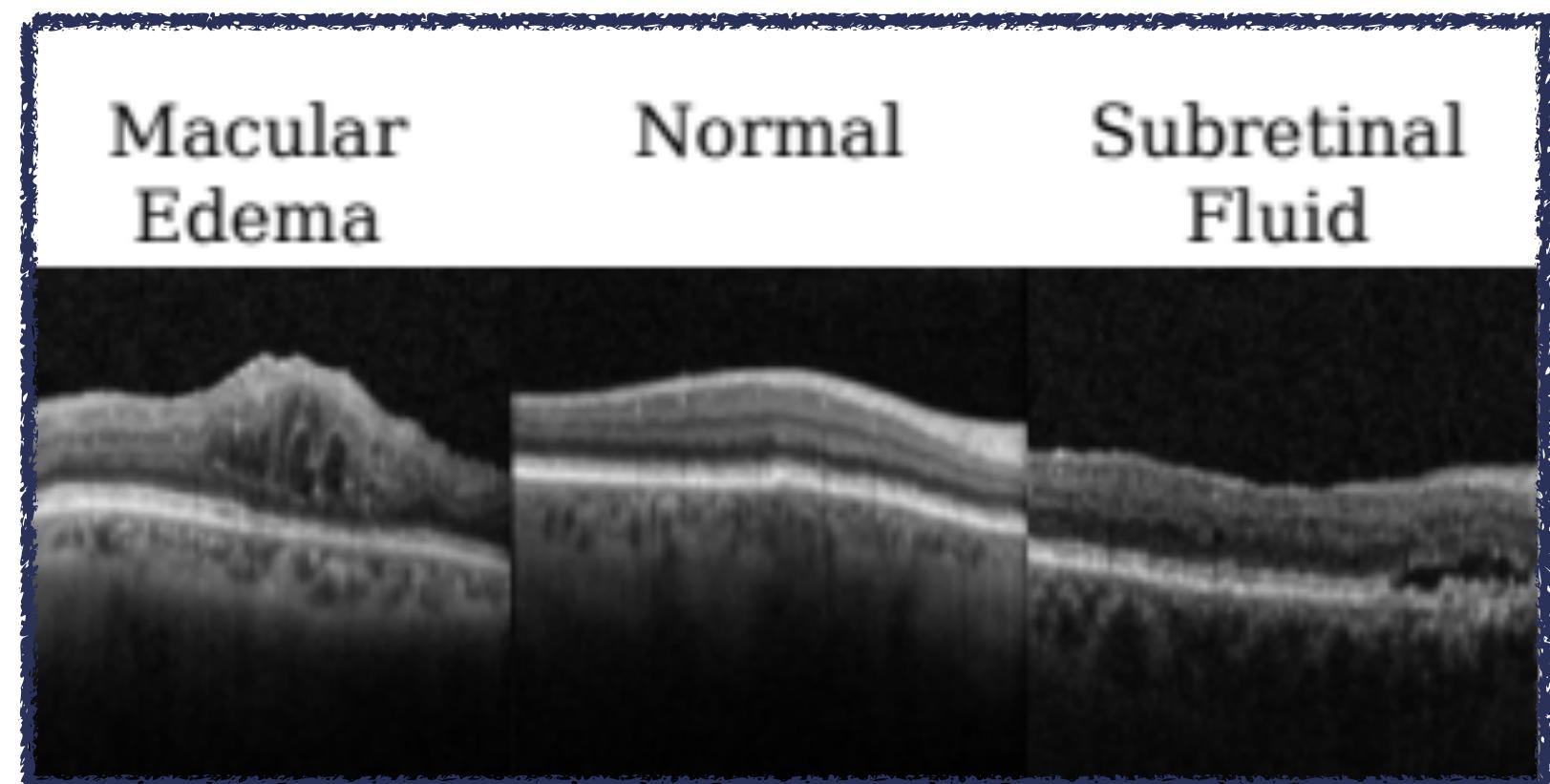
## データセット

- 3つのデータセットを用いて有効性を確認していく。

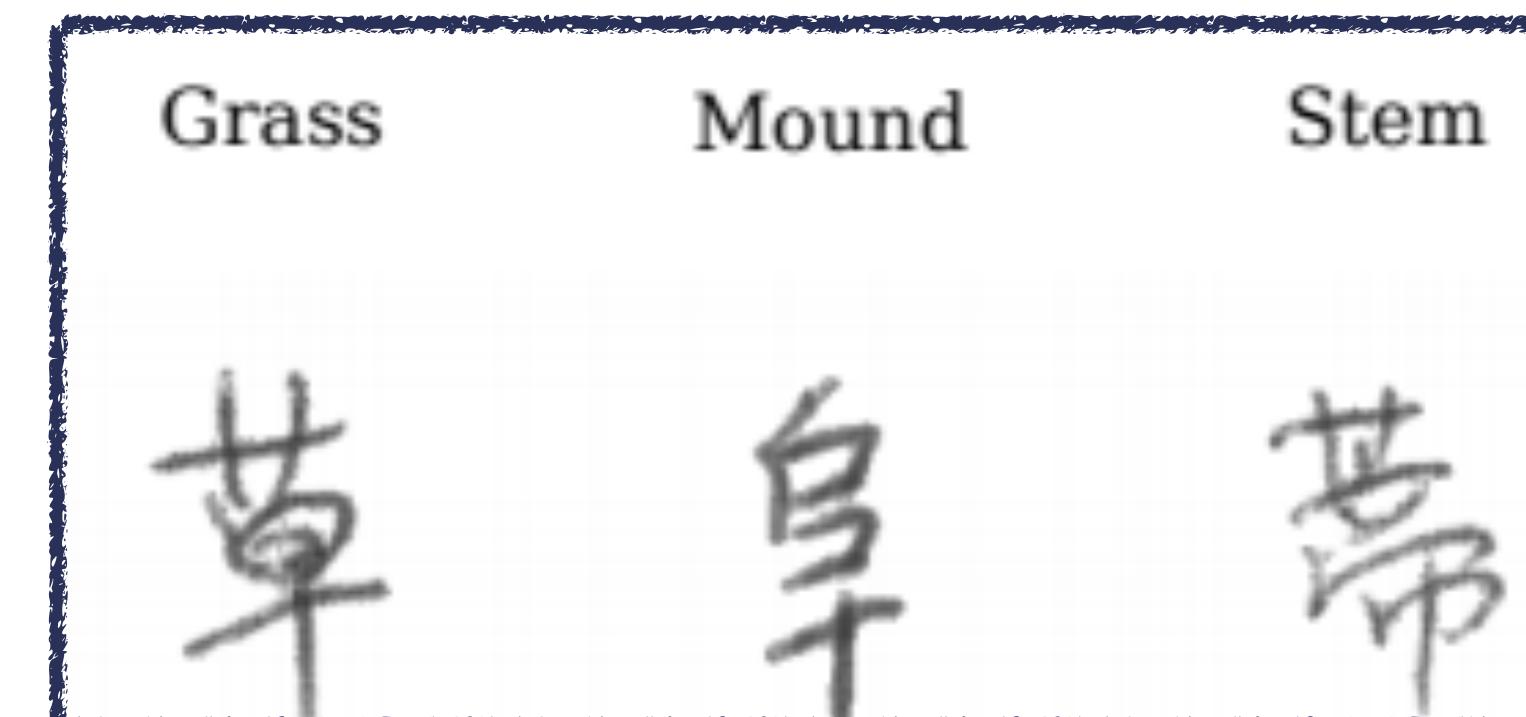
1. Butterflies (蝶識別)



2. OCT Eyes (網膜診断)



3. Chinese Characters (文字識別)



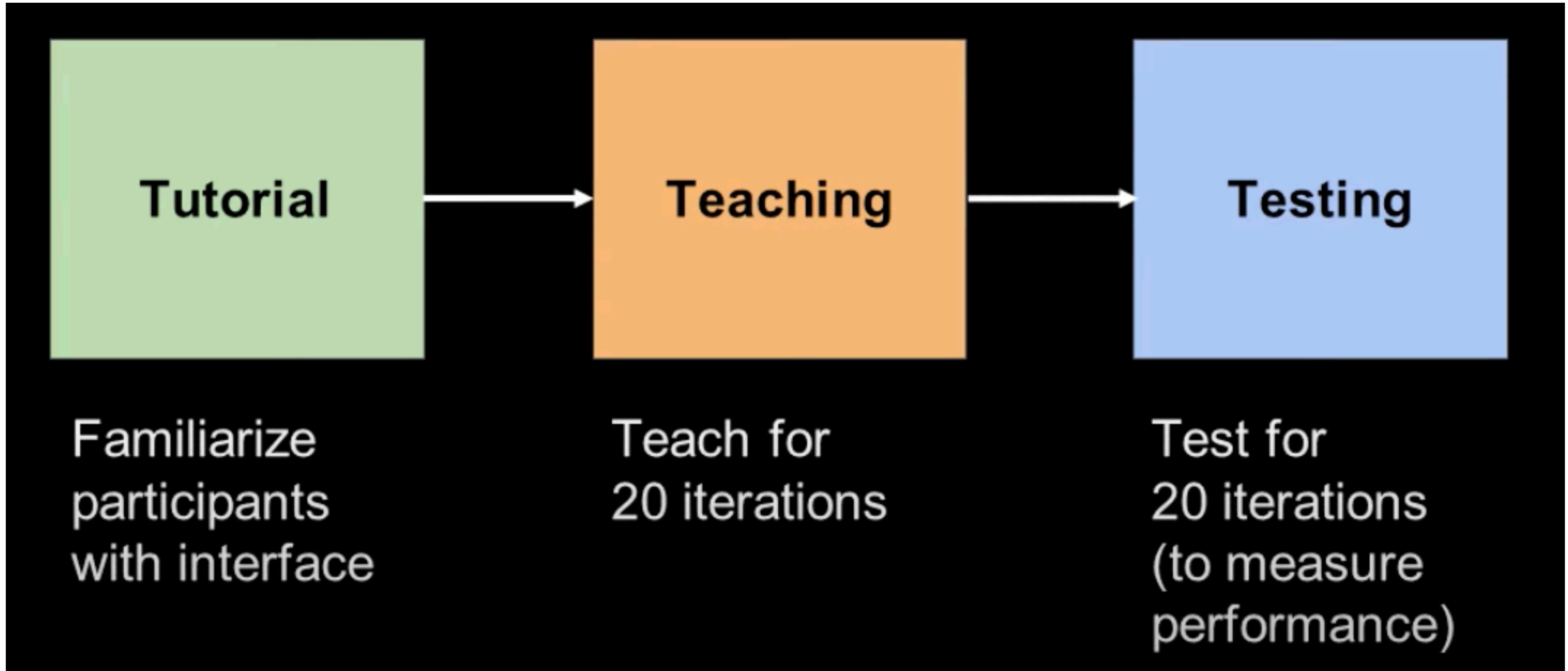
# 実験設定

## ちょっと詳細に

- Amazon Mechanical Turkで被験者40人
- 模試画像はランダムに提示して、選択回答の順番もランダムに
  - 位置によるバイアスをなくしている。
- 対抗手法
  - RAND\_IM: ランダムに画像と正解ラベル
  - RAND\_EXP: ランダム画像とその説明
  - STRICT: いい感じの画像を選択する

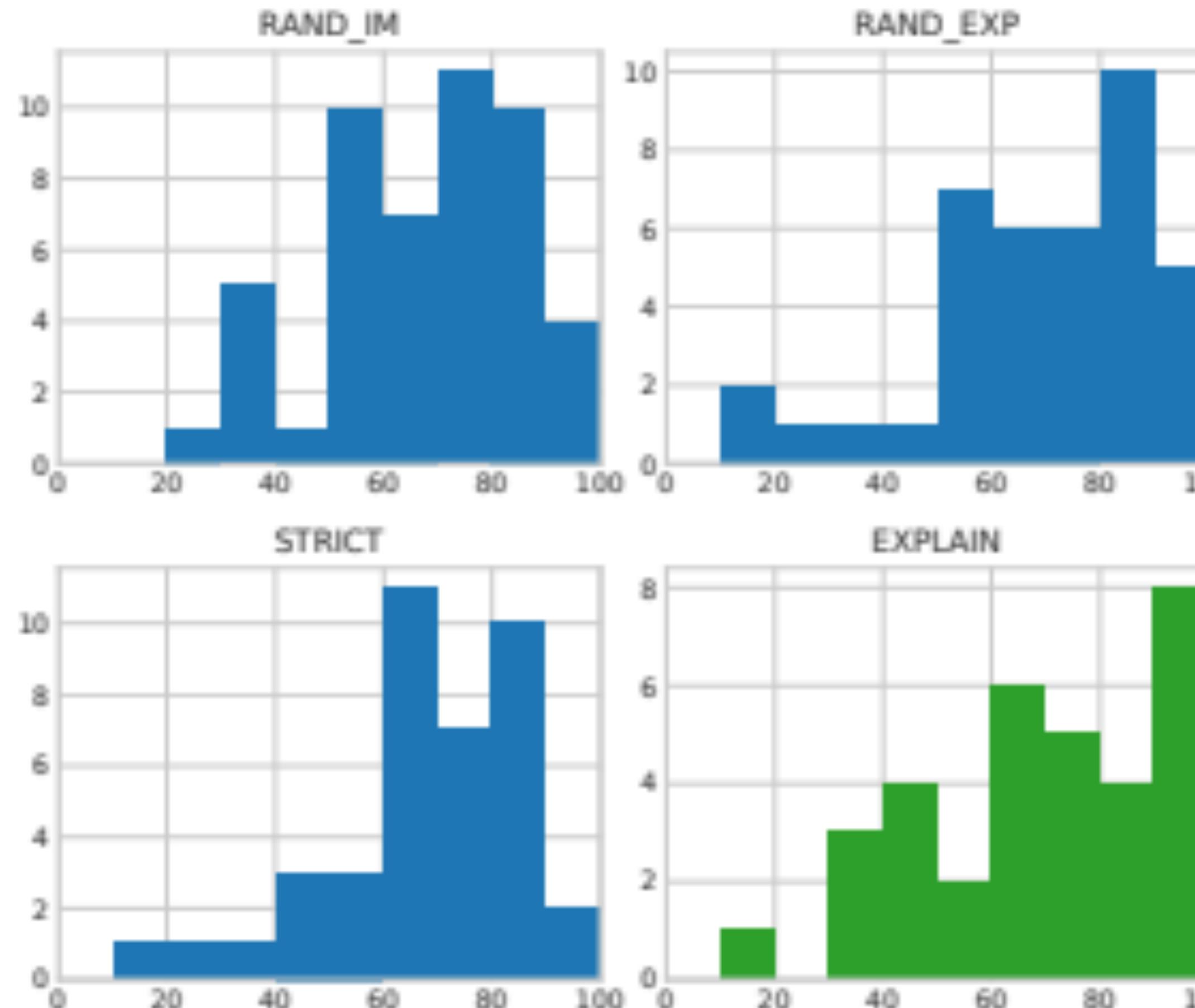
# 実験設定

## 機械教示の流れ



# 実験結果

## Butterfly



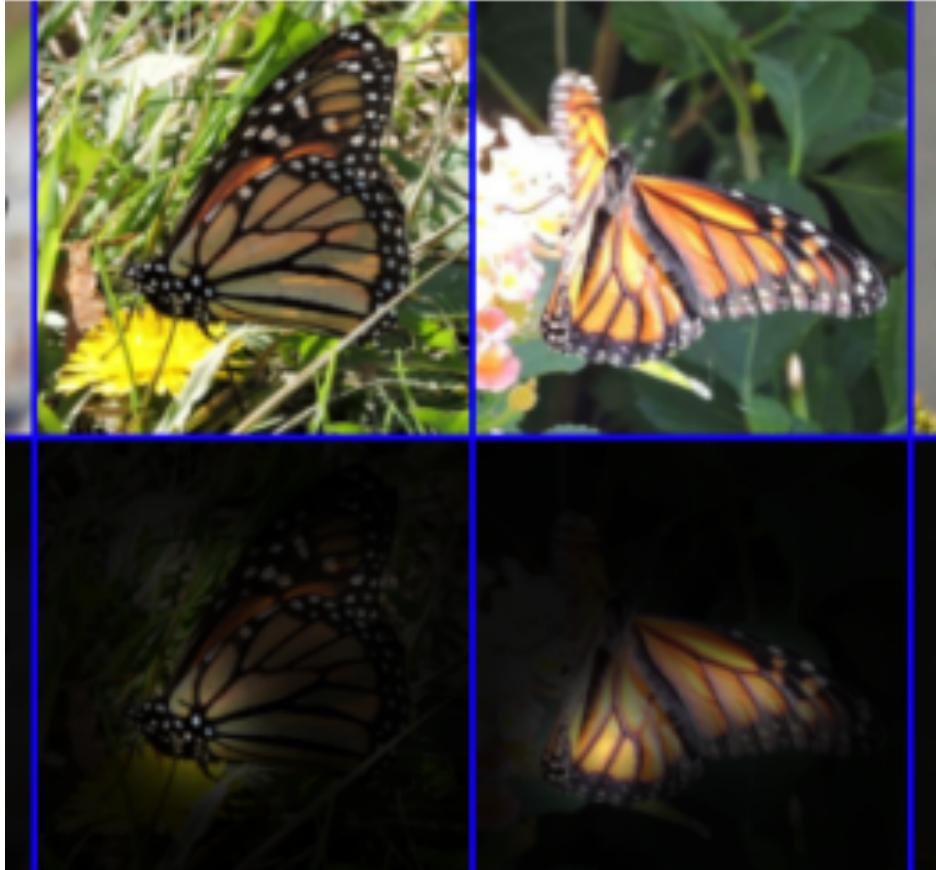
このデータセットは難しい傾向  
似た3種は混乱しがち

正答率のヒストグラム  
が右にシフトしている

A) Butterflies

# 実験

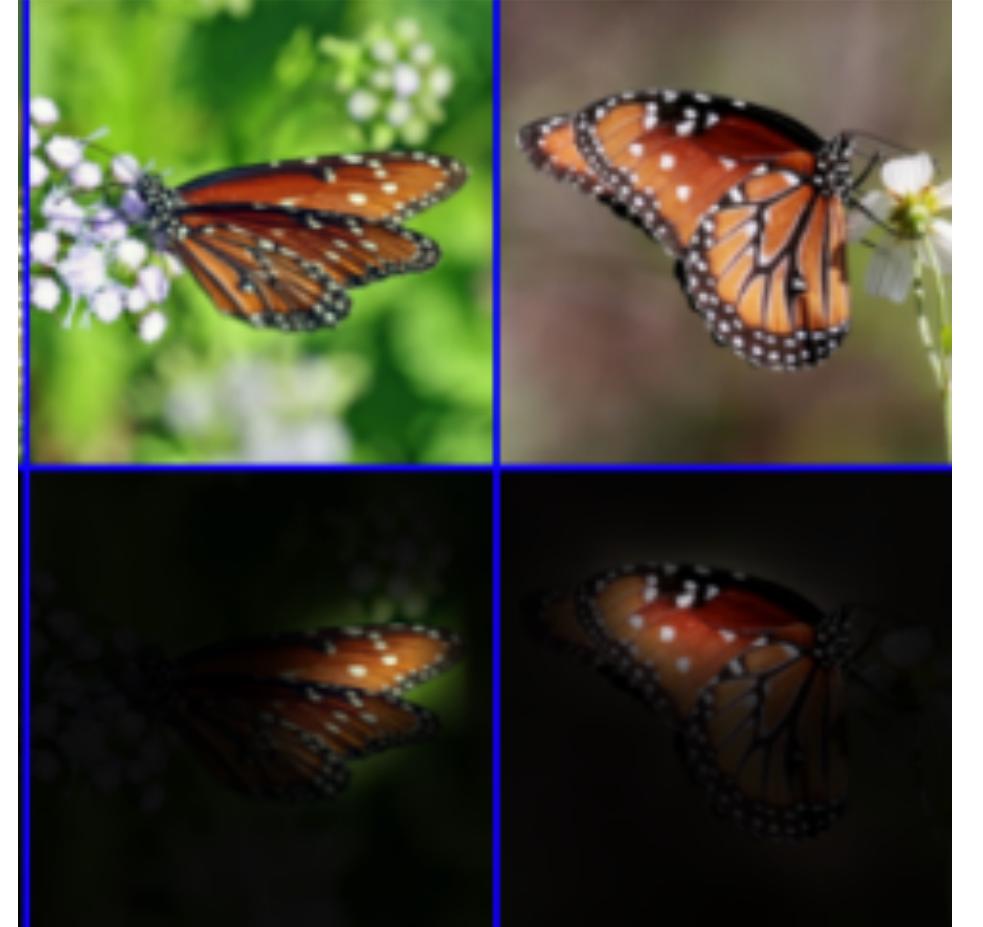
## 説明画像のイメージ: Butterfly



A) Monarch



B) Viceroy



C) Queen



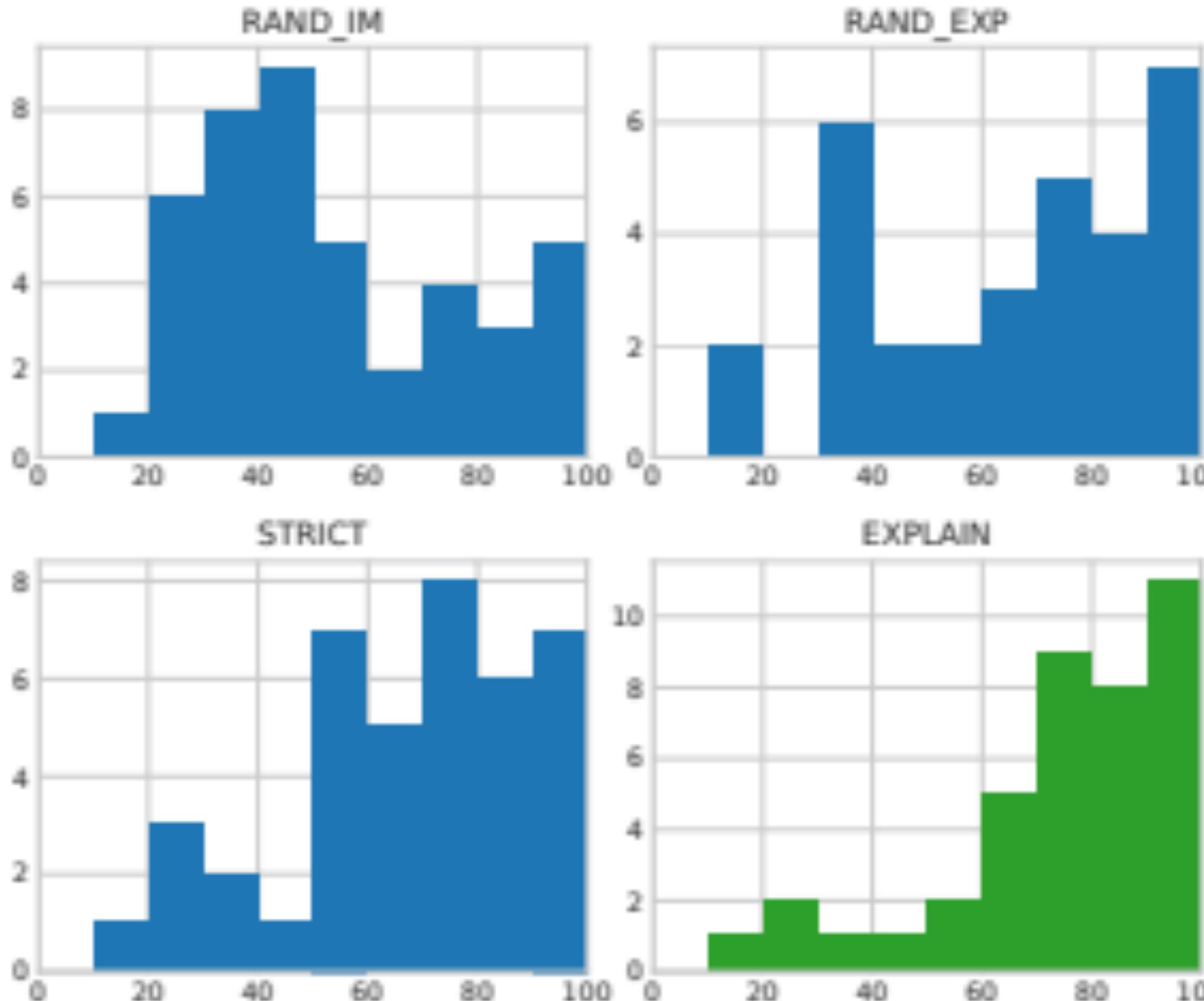
D) Red Admiral



E) Cabbage White

# 実験結果

## OCT Eyes



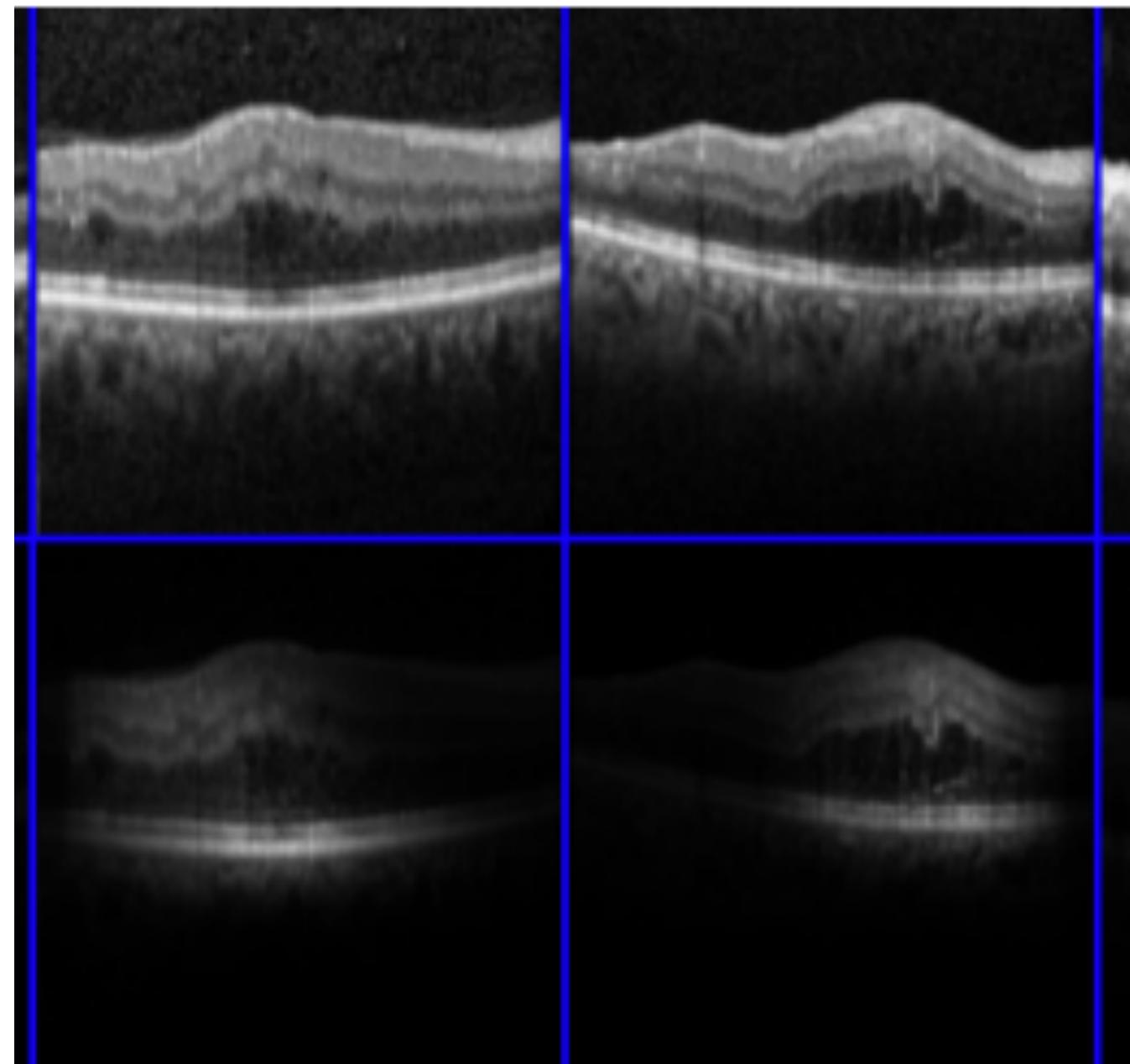
ランダムでも向上  
してしまって

正答率のヒストグラム  
が右にシフトしている

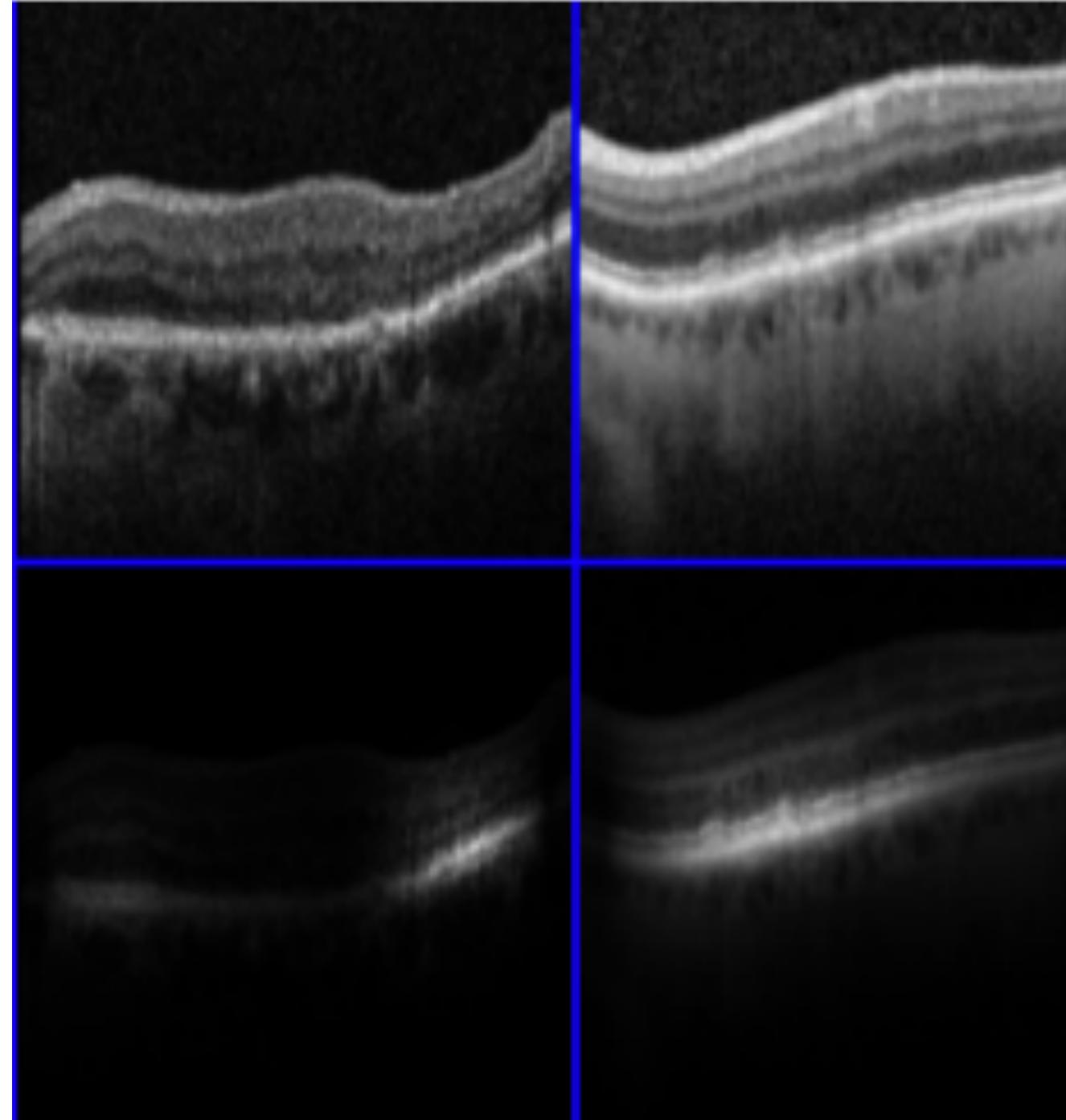
B) OCT Eyes

# 実験

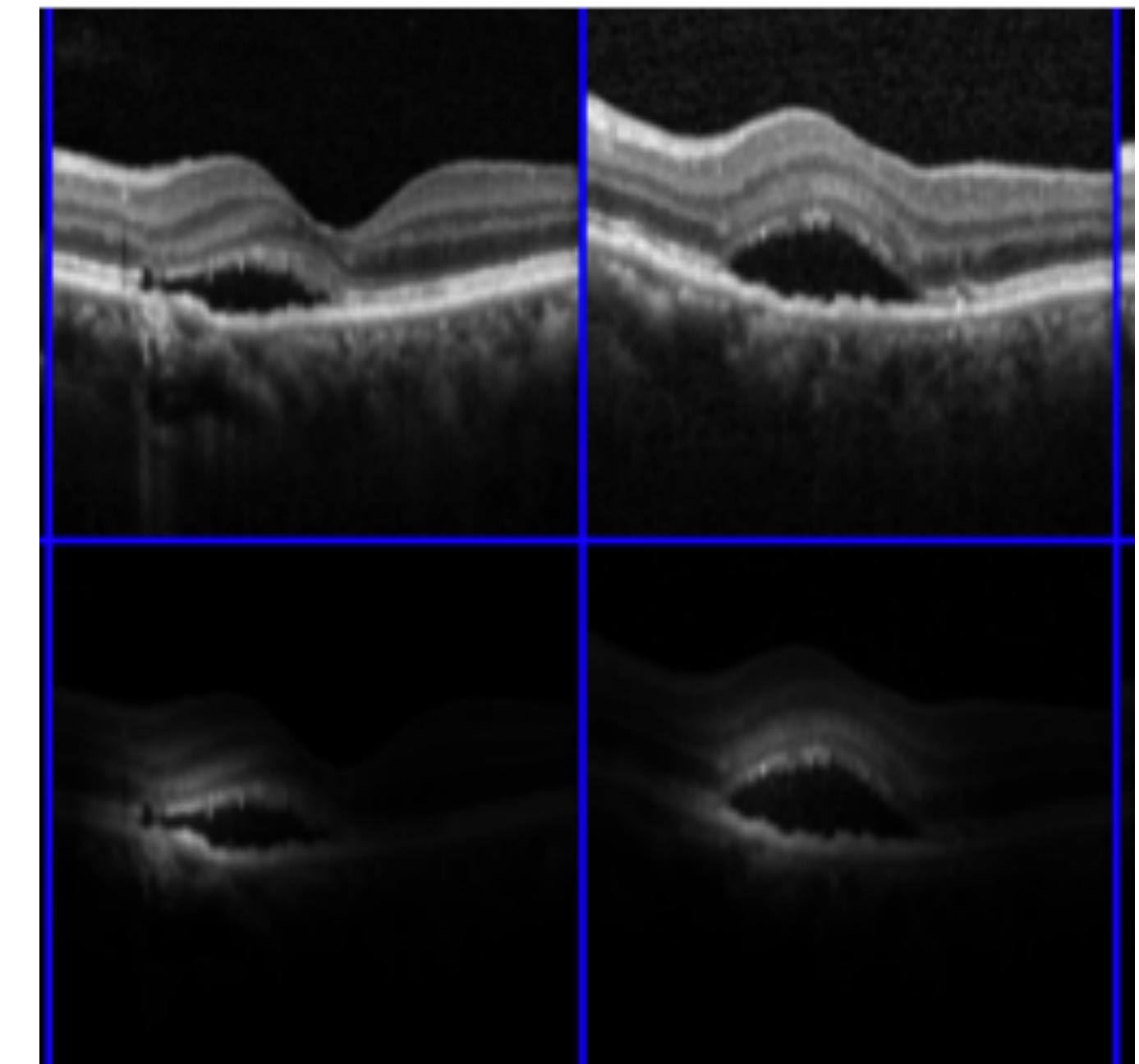
## 説明画像のイメージ: OCT Eyes



A) Macular Edema



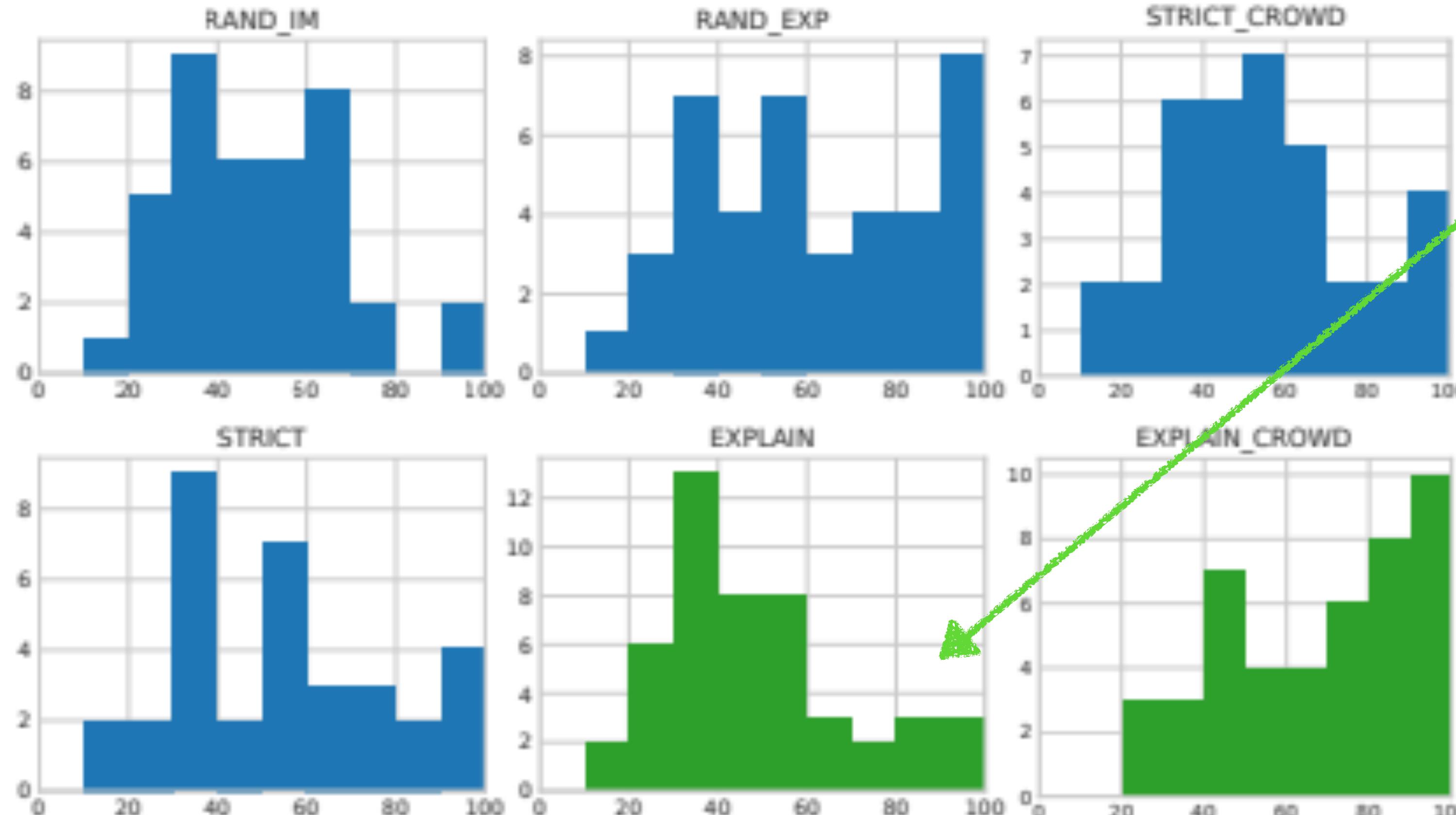
B) Normal



C) Subretinal Fluid

# 実験結果

## Chinese Character



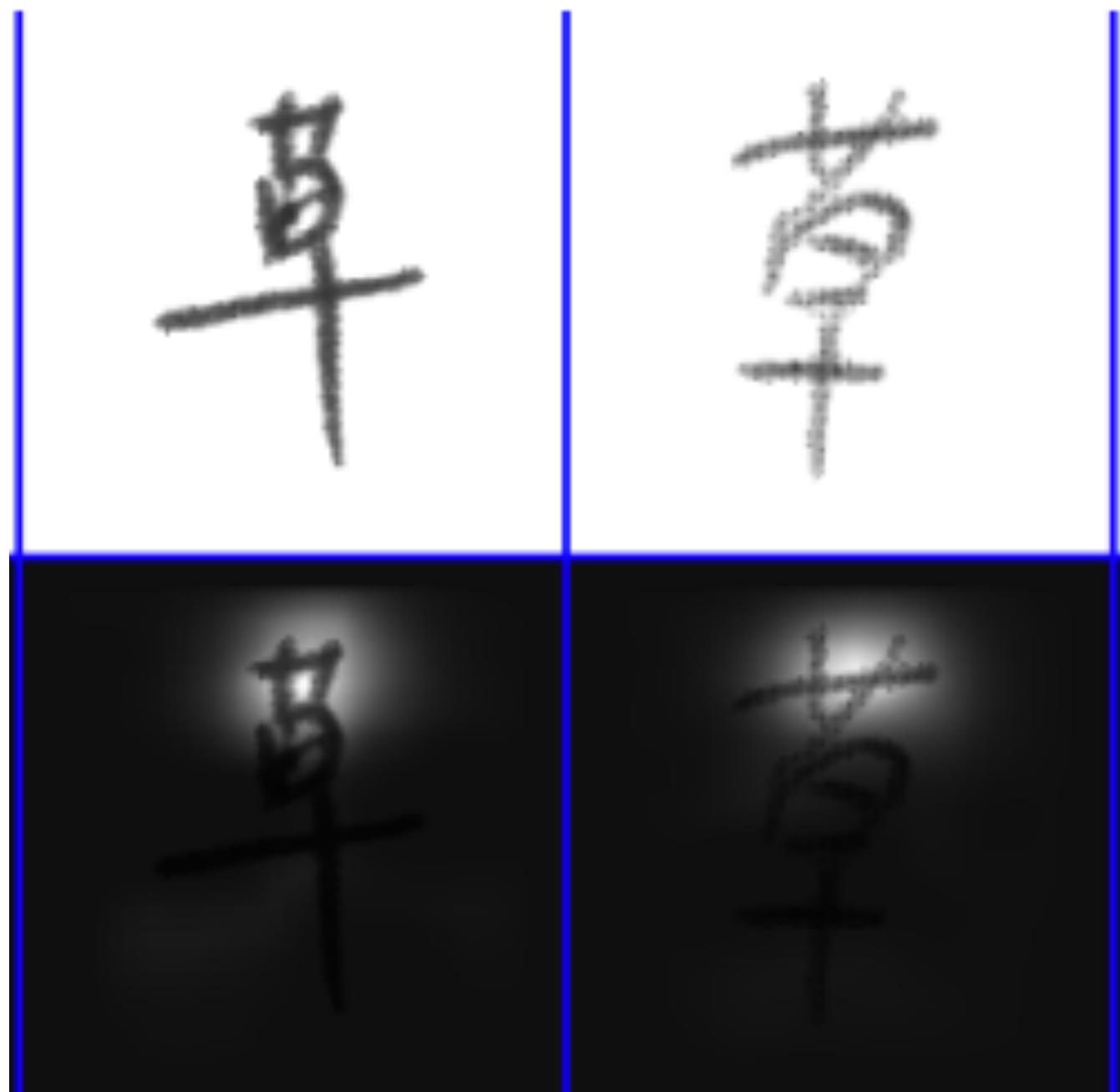
CNNの説明が失敗している

手動の説明が性能良い

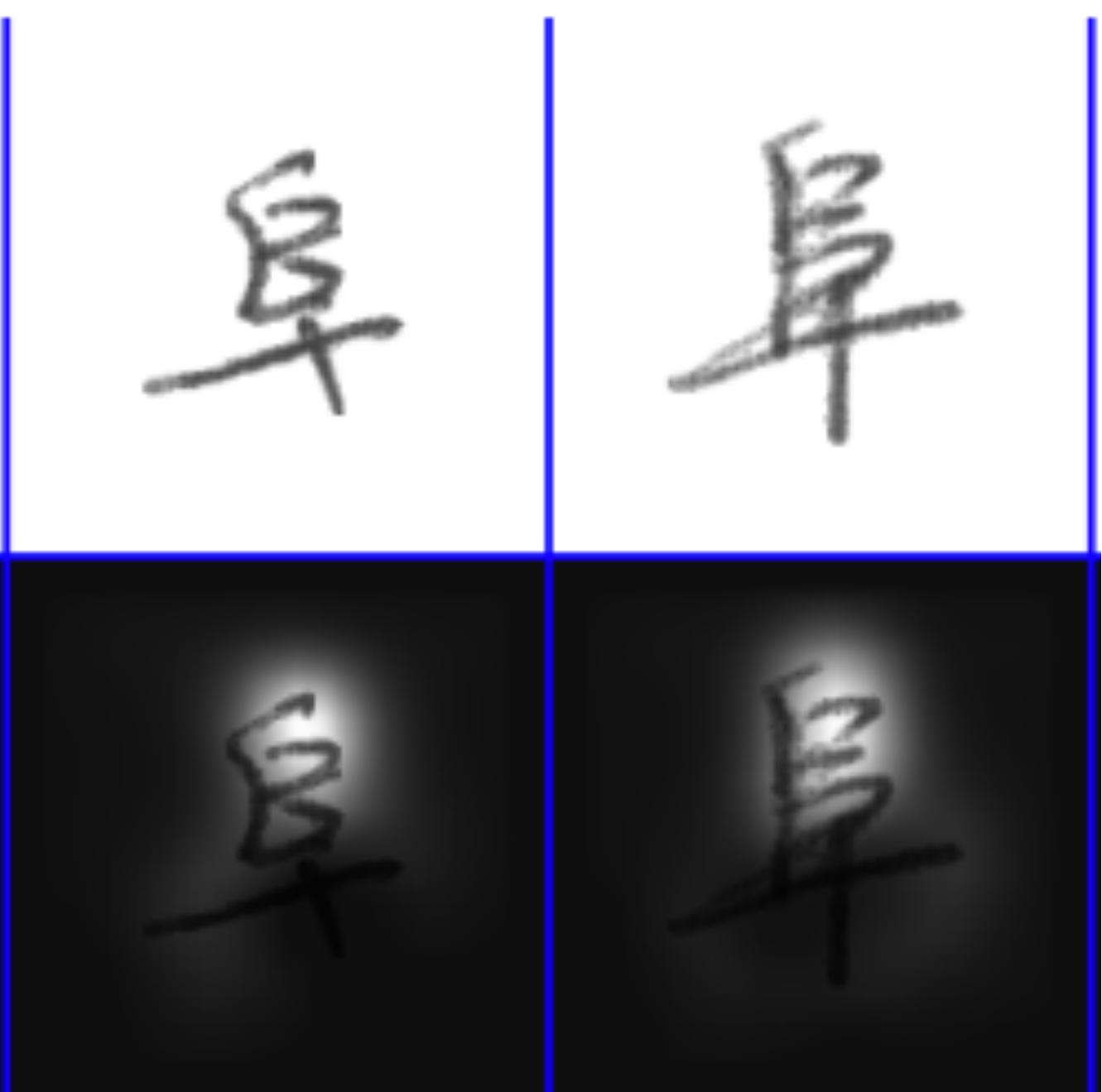
C) Chinese Characters

# 実験

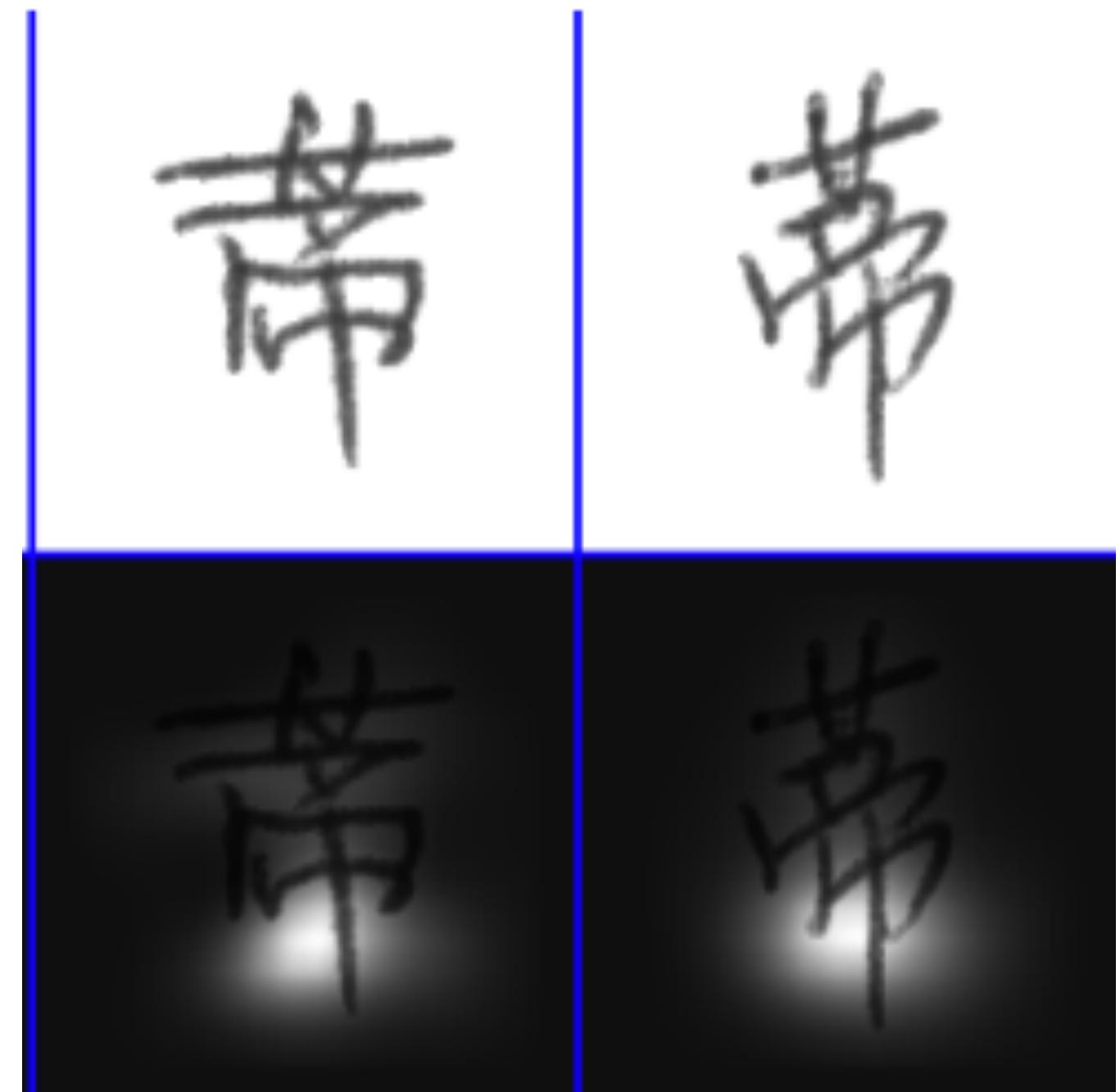
説明画像のイメージ: Chinese Character



A) Grass



B) Mound



C) Stem

# まとめ

## 画像説明付き機械教示

- 視覚的説明性を画像に与えてそれをもとに機械教示を行なっていく。
- 従来の正解ラベルだけ教える方法より、説明がある方が学習効果が高く、さらに効果の高い教材集合を見つけていている。
- 将来的にオンラインでインタラクティブにやりたいよね

# まとめ

やっぱりやられてた

- 蝶の識別ぐらい細かいやつが自分としては教育したい問題設定
- 目の画像は専門知識なくても違いが分かれば識別可能は魅力的
- 漢字識別はCNNの注目点でも良さそうなのにどこがダメなんだ  
ろう、人力での説明生成も載せて欲しかった。