Artist Agent:
A Reinforcement Learning Approach to Automatic Stroke Generation in Oriental Ink Painting 紹介

望月 駿一

本論文のモチベーション

- ✓ 筆を使った自然な墨絵タッチの絵を生成したい
- ✓ 墨絵とは…
  - ✓ 油絵等のタッチを重ねる西洋画と異なり、わずかな筆跡で情景内の特徴的な情報を表現する技法
  - ✓ 筆跡は、描く対象の形状、筆の姿勢と軌跡、筆に含まれる墨の分によって決定される
  - ✓ いま(2012年)世界中で最も注目されているペインティングスタイルのひとつ（？）
- ✓ 墨絵を描く難しさ
  - ✓ 原画が持つ複雑な映像情報を抽象的に表現する必要がある
  - ✓ 滑らかで自然な筆運びを再現する必要がある

著者(Xin)らの先行研究（DP）

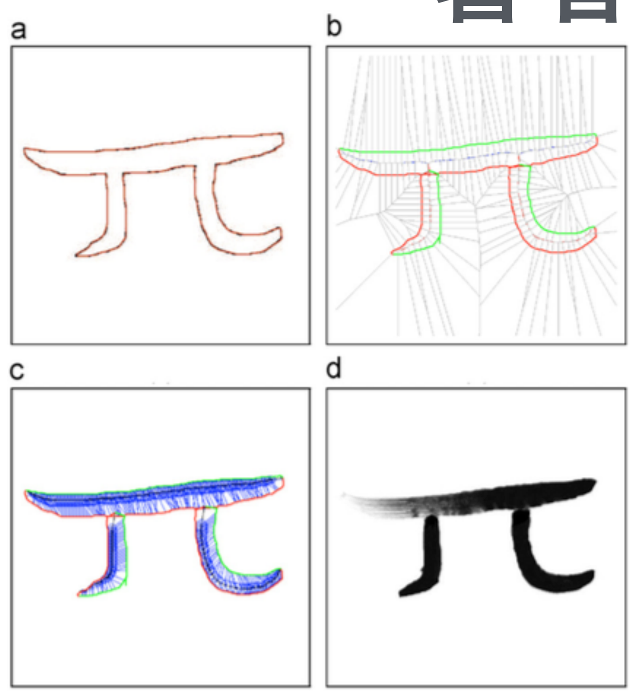
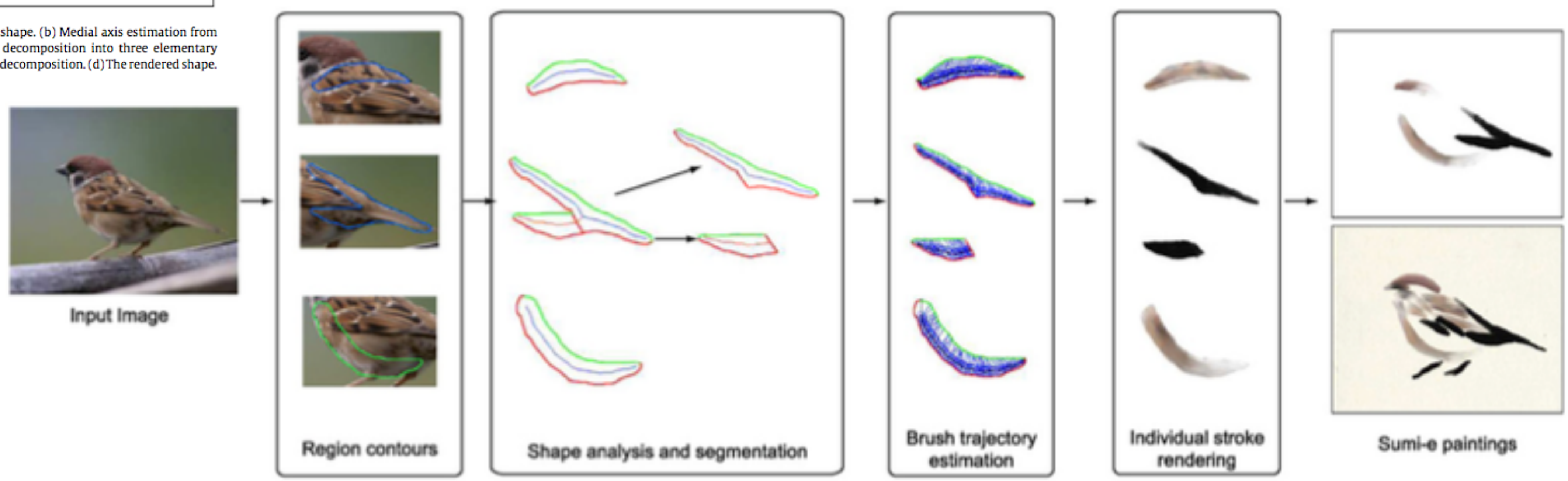


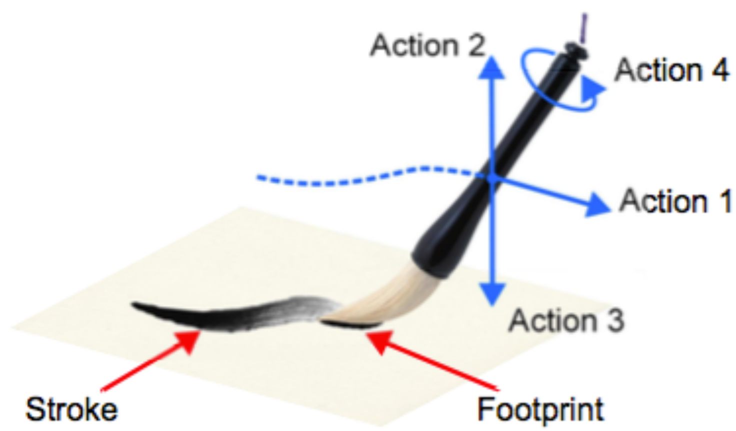
Fig. 8. Handling complex shapes: (a) Input shape. (b) Medial axis estimation from the Voronoi diagram of the shape and its decomposition into three elementary strokes. (c) Brush trajectory estimation after decomposition. (d) The rendered shape.



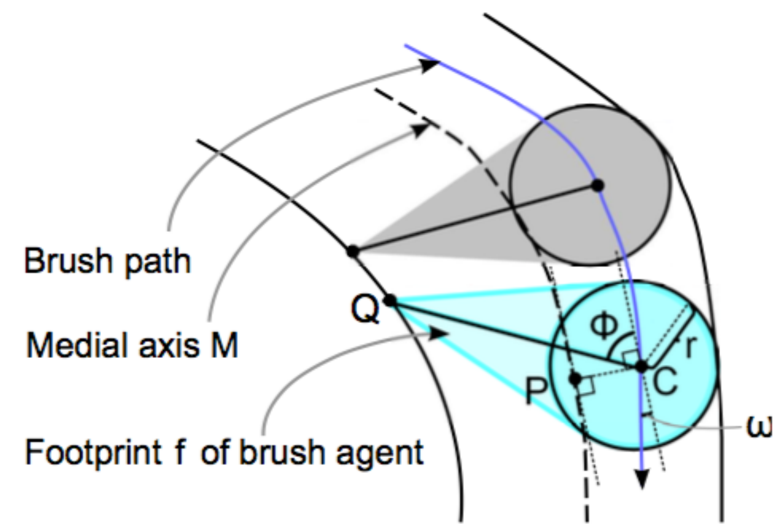
複雑な図形の分割を自分たちで行う必要がある

提案手法の概要

- 筆を筆先と輪郭に関係を表す状態と筆先を動かすアクションをもつ強化学習のエージェントと見立てる

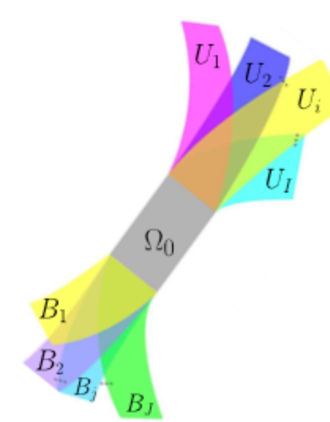


(a) Brush model



- 塗りを構成する基本図形について学習させてエピソードの組み合わせで複雑な図形を表現。輪郭は用意する。

(c)



(a)

$(\mathcal{S}, \mathcal{A}, p_{\mathrm{I}}, p_{\mathrm{T}}, R)$ からなるMDPを考える

$\mathcal{S}$ ：連続な状態の集合（後ほど詳しく定義）

$\mathcal{A}$ ：連続なアクションの集合（後ほど詳しく定義）

p_{I} ：初期状態の確率分布

$p_{\mathrm{T}}(\boldsymbol{s}'|\boldsymbol{s}, a)$ ：アクションaに対する状態遷移確率

$R(\boldsymbol{s}, a, \boldsymbol{s}')$ ：sがaによりs'に遷移したときの即時報酬

$\pi(a|\boldsymbol{s}; \boldsymbol{\theta})$ ：状態s∈Sにおける行動aの条件つき確率を表現する確率的方策（パラメタθに依存）

報酬の設計

長さTのトラジェクトリーを以下のhとして

$$h \quad = \quad (\boldsymbol{s}_1, a_1, \ldots, \boldsymbol{s}_T, a_T, \boldsymbol{s}_{T+1})$$

hの累積割引報酬を以下と定義する

$$R(h) = \sum_{t=1}^{T} \gamma^{t-1} R(\boldsymbol{s}_t, a_t, \boldsymbol{s}_{t+1})$$

ただし$\gamma \in [0,1)$を割引率とする

目的関数の設計

期待報酬J(θ)を最大化する θ *を見つけて
最良の方策p(h| θ *)を決定したい

$$J(\boldsymbol{\theta}) = \int p(h|\boldsymbol{\theta})R(h)\mathrm{d}h$$

$$\boldsymbol{\theta}^* \equiv \mathrm{argmax}\ J(\boldsymbol{\theta})$$

ただしp(θ)のもとでトラジェクトリーhとなる確率は以下で表現される

$$p(h|\boldsymbol{\theta}) = p(\boldsymbol{s}_1)\prod_{t=1}^{T} p(\boldsymbol{s}_{t+1}|\boldsymbol{s}_t, a_t)\pi(a_t|\boldsymbol{s}_t, \boldsymbol{\theta})$$

方策勾配法

θ*を発見するために方策勾配法を用いる

$$\boldsymbol{\theta} \longleftarrow \boldsymbol{\theta} + \varepsilon \nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta})$$

ただし∇J(θ)は以下で与えられる

$$\begin{aligned}
\nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta}) &= \int \nabla_{\boldsymbol{\theta}} p(h|\boldsymbol{\theta}) R(h) \mathrm{d}h \\
&= \int p(h|\boldsymbol{\theta}) \nabla_{\boldsymbol{\theta}} \log p(h|\boldsymbol{\theta}) R(h) \mathrm{d}h \\
&= \int p(h|\boldsymbol{\theta}) \sum_{t=1}^{T} \nabla_{\boldsymbol{\theta}} \log \pi(a_t|\boldsymbol{s_t}, \boldsymbol{\theta}) R(h) \mathrm{d}h
\end{aligned}$$

p(h|θ)の表現

$$\nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta}) \quad = \int p(h|\boldsymbol{\theta}) \sum_{t=1}^{T} \nabla_{\boldsymbol{\theta}} \log \pi(a_t|\boldsymbol{s_t}, \boldsymbol{\theta}) R(h) \mathrm{d}h$$

方策に対する真のトラジェクトリー分布p(h|θ)
は分からないので経験分布による近似を用いる

$$\nabla_{\boldsymbol{\theta}} \widehat{J}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{n=1}^{N} \sum_{t=1}^{T} \nabla_{\boldsymbol{\theta}} \log \pi(a_t^{(n)}|\boldsymbol{s}_t^{(n)}, \boldsymbol{\theta}) R(h^{(n)})$$

where $\{h^{(n)}\}_{n=1}^{N}$ are N episodic samples with T steps and $h^{(n)} = (\boldsymbol{s}_1^{(n)}, a_1^{(n)}, \ldots, \boldsymbol{s}_T^{(n)}, a_T^{(n)}, \boldsymbol{s}_{T+1}^{(n)})$.

方策 $\pi(a|s;\theta)$ の設計

ガウス分布を方策に採用

Let us employ the Gaussian policy function with parameter $\boldsymbol{\theta} = (\boldsymbol{\mu}^\top, \sigma)^\top$, where $\boldsymbol{\mu}$ is the mean vector and σ is the standard deviation:

$$\pi(a|\boldsymbol{s};\boldsymbol{\theta}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(a - \boldsymbol{\mu}^\top \boldsymbol{s})^2}{2\sigma^2}\right).$$

ガウス方策分布における期待報酬の導出

Then the derivatives of the expected return $J(\boldsymbol{\theta})$ with respect to the parameter $\boldsymbol{\theta}$ are given as

$$\nabla_{\boldsymbol{\mu}} \log\pi(a|\boldsymbol{s};\boldsymbol{\theta}) = \frac{a-\boldsymbol{\mu}^\top \boldsymbol{s}}{\sigma^2}\boldsymbol{s},$$

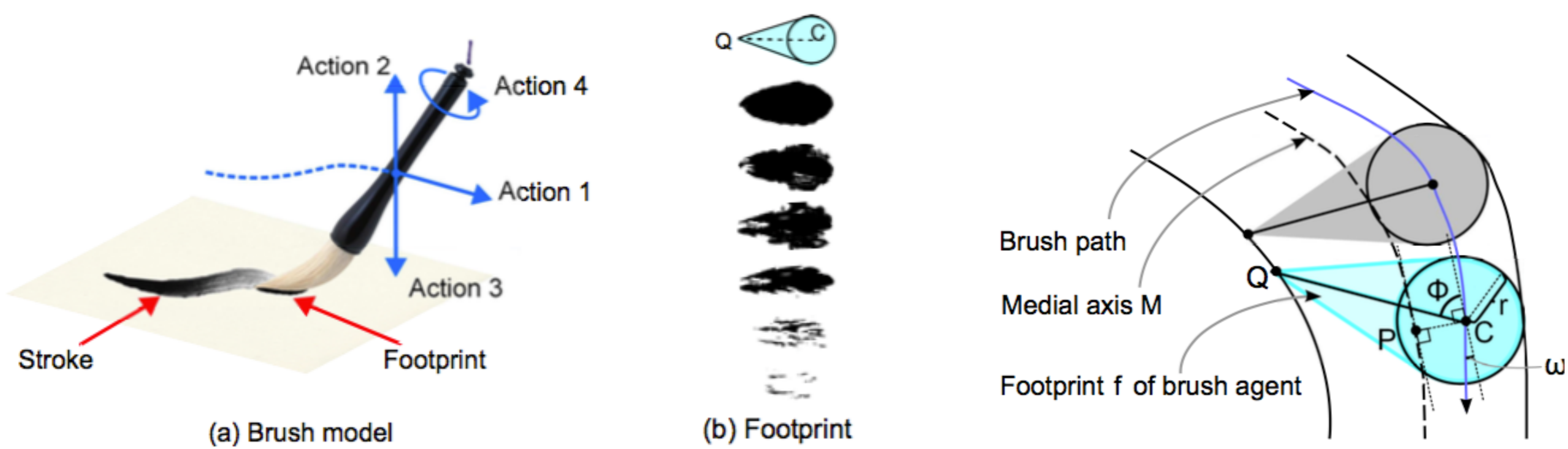
$$\nabla_{\sigma} \log\pi(a|\boldsymbol{s};\boldsymbol{\theta}) = \frac{(a-\boldsymbol{\mu}^\top \boldsymbol{s})^2-\sigma^2}{\sigma^3}.$$

$$\nabla_{\boldsymbol{\mu}} J(\boldsymbol{\theta}) = \frac{1}{N}\sum_{n=1}^{N}(R(h^{(n)})-b)\sum_{t=1}^{T}\frac{(a_t^{(n)}-\boldsymbol{\mu}^\top \boldsymbol{s}_t^{(n)})\boldsymbol{s}_t^{(n)}}{\sigma^2},$$

$$\nabla_{\sigma} J(\boldsymbol{\theta}) = \frac{1}{N}\sum_{n=1}^{N}(R(h^{(n)})-b)\sum_{t=1}^{T}\frac{\left(a_t^{(n)}-\boldsymbol{\mu}^\top \boldsymbol{s}_t^{(n)}\right)^2-\sigma^2}{\sigma^3},$$

筆エージェントの設計

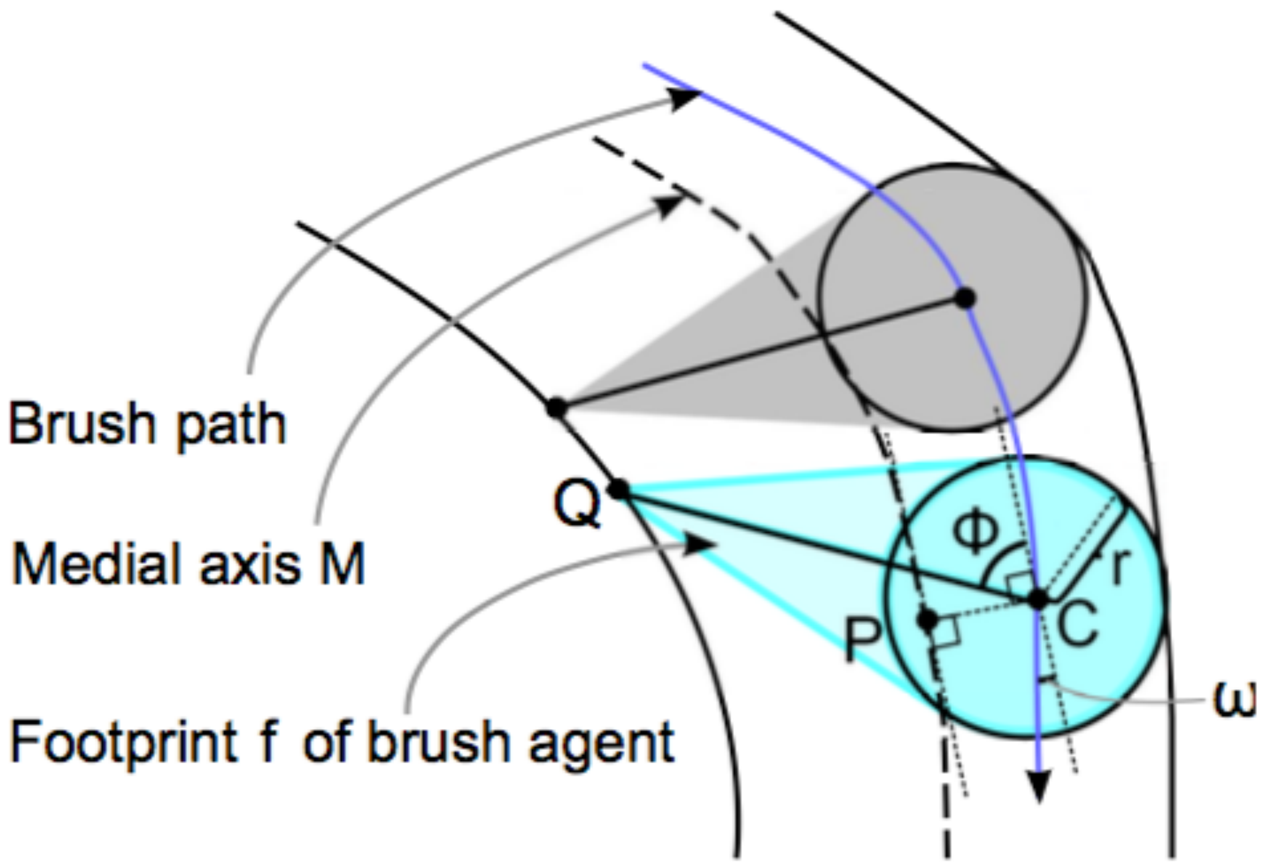
筆エージェントの行動の設計



- 墨絵には本来、筆を①並行移動②上げる③下げる④軸を回転する動きがある
- ①の動きが決まれば、残りは頂点と円の接線が輪郭に接するように最適化すると仮定

筆エージェントの状態の設計

現在の筆跡fを頂点Qから円Cへの接線からなる図形とする



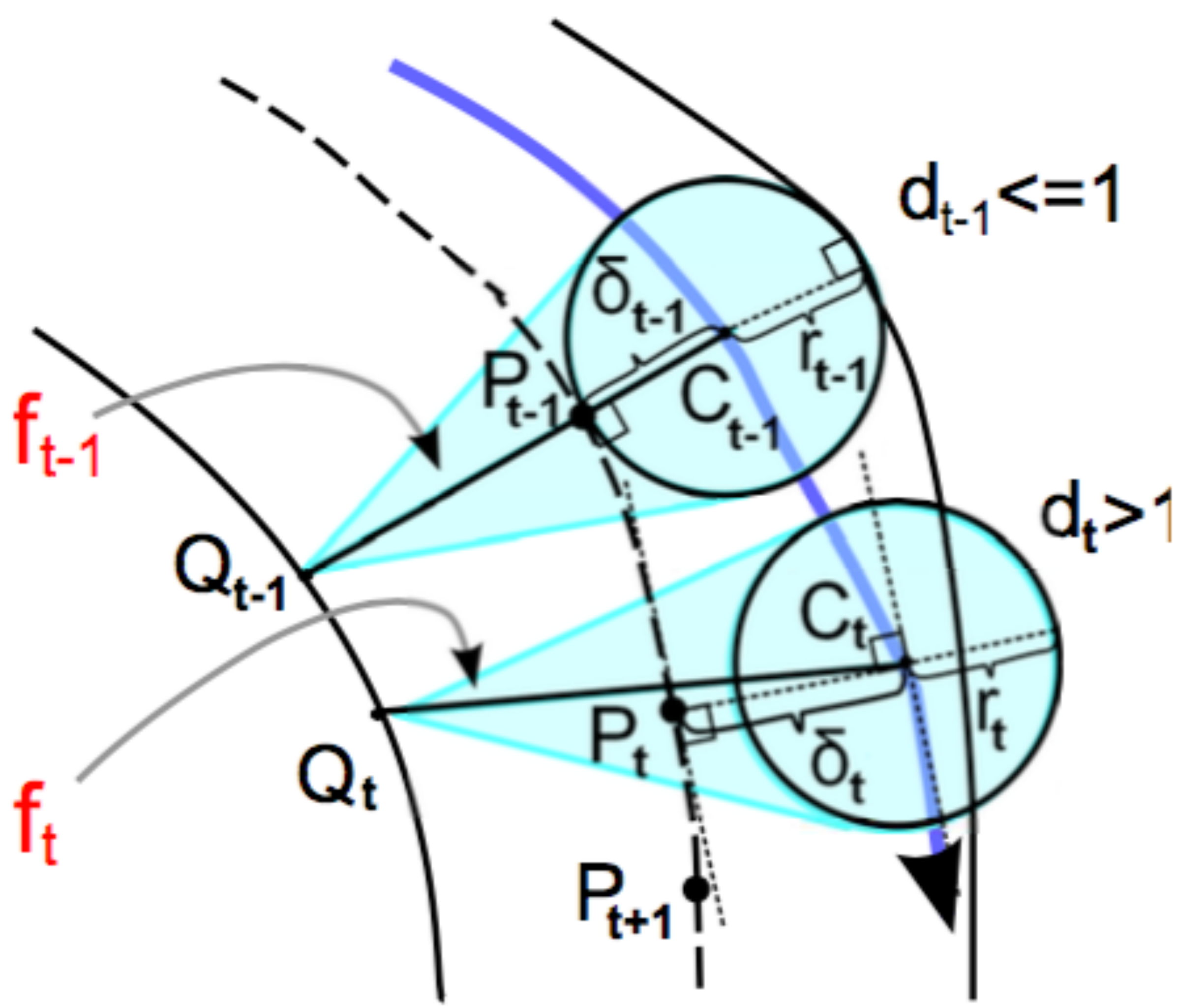
M：輪郭から計算した中間線

$\omega \in (-\pi, \pi]$
中間線の接線と進行方向の相対角度

$\phi \in (-\pi, \pi]$
中間線の接線と円中心から頂点へ引いた線の相対角度

$l \in \{0, 1\}$
任意の状態が既にfによってカバーされた領域を含むか判定する変数（0ならカバー済み）

筆エージェントの報酬の設計①



$d \in [-2, 2]$

- Pから中間線への垂線距離をδ、円半径をrとした時、|d|=δ/rで表現される値
- 筆跡の円部分が中間線と輪郭の間でどのような関係になっているかを示す値
- d>0で右側、d<0で左側境界を円が含んでいる状態
- |d|>1で境界を円がはみ出した状態。円中心は境界内という定義なので|d|<2

$\kappa_i (i = 1, 2) \in [0, 1)$

$$|\kappa_i| = \frac{2}{\pi} \arctan\left(\frac{\alpha}{\sqrt{r'_i}}\right)$$

- 輪郭の半径r'_iに対してαをパラメタとしてカーブの急峻さ(=塗りの難しさ)を表現する変数
- i=1は現在、i=2は次の状態で接するカーブに対する急峻さ表現

筆エージェントの報酬の設計②

より自然な良い筆運びを実現する動きに高い報酬を与えたい

$$R(\boldsymbol{s}_t, a_t, \boldsymbol{s}_{t+1}) = \begin{cases} 0 & \text{if } f_t = f_{t+1} \text{ or } l = 0, \\ \dfrac{1 + (|\kappa_1(t)| + |\kappa_2(t)|)/2}{\lambda_1 E^{(t)}_{\text{location}} + \lambda_2 E^{(t)}_{\text{posture}}} & \text{otherwise.} \end{cases}$$

輪郭に対する塗りの良さを測る指標E_loc

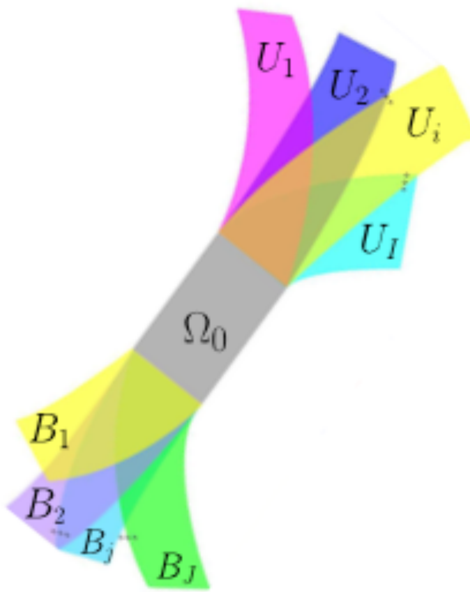
$$E^{(t)}_{\text{location}} = \begin{cases} \tau_1 |\omega_t| + \tau_2(|d_t| + W) & d \in [-2, -1) \cup (1, 2], \\ \tau_1 |\omega_t| & d \in [-1, 1], \end{cases}$$

自然な運筆かどうかを測る指標E_pos（大きいと不自然）

$$E^{(t)}_{\text{posture}} = \zeta_1 \Delta\omega_t + \zeta_2 \Delta\phi_t + \zeta_3 \Delta d_t,$$

$$\Delta x_t = \begin{cases} 1 & \text{if } x_t = x_{t-1} = 0, \\ \dfrac{(x_t - x_{t-1})^2}{(|x_t| + |x_{t-1}|)^2} & \text{otherwise.} \end{cases}$$

筆エージェントの学習方法の設計

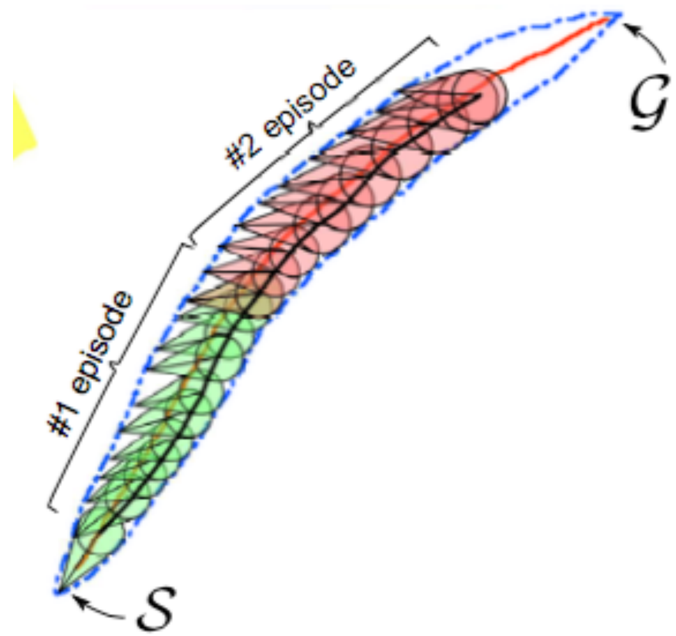


(a)


(c)

- 複雑な図形もいくつかの基本的な構造（とめ、はねはらいなど）から成り立っていると仮定
- (a)のように新規の図形が来ても組み合わせで対応が可能
- 80の基本的構造の手書きライブラリを用意して最初に学習させた

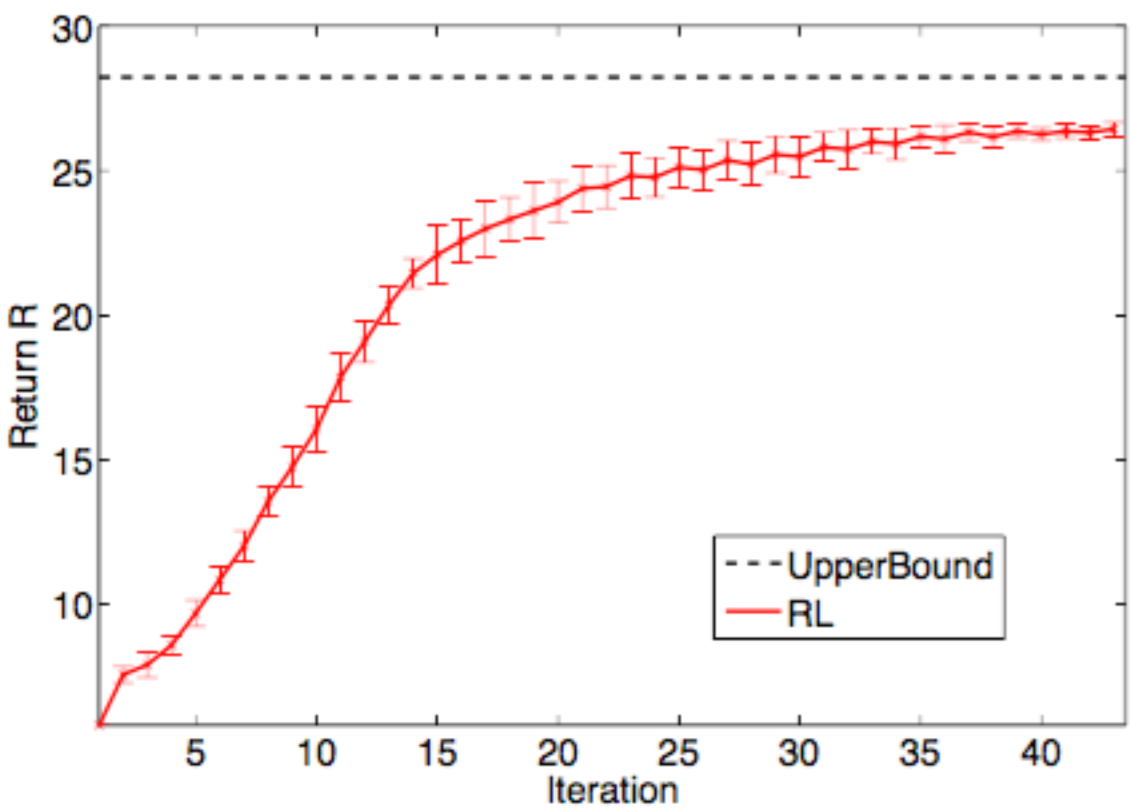
筆エージェントの学習方法の設計

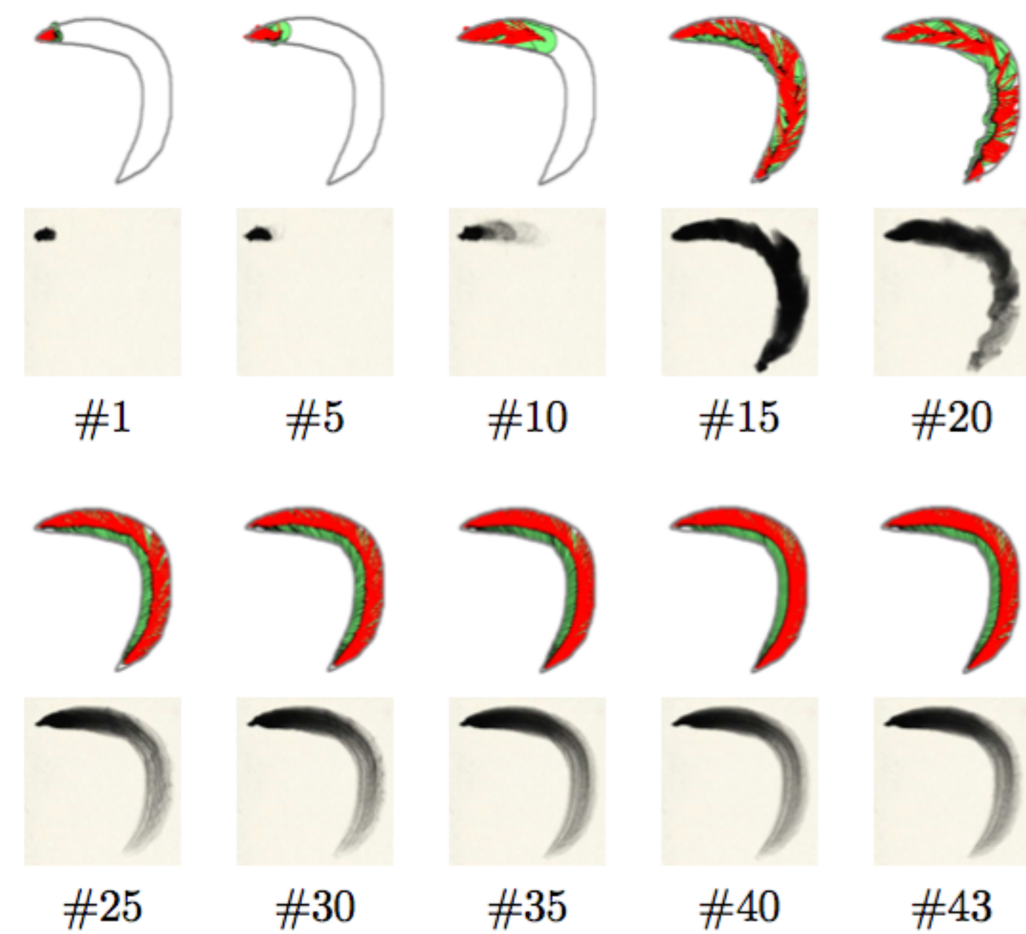


(b)

- 一つの図形を分割したエピソードhが正しい順に並ぶ
  $H = [h^{(1)}, h^{(2)}, \ldots, h^{(N)}]$を用意した
- 最初のエピソードは必ず中間軸上のSから始まるように設定
- 次回以降のエピソードは前回の始点か終点付近からスタート
- 塗り余地がない場合と、残りのトラジェクトリーがTステップより少ない場合は前回始点に戻る

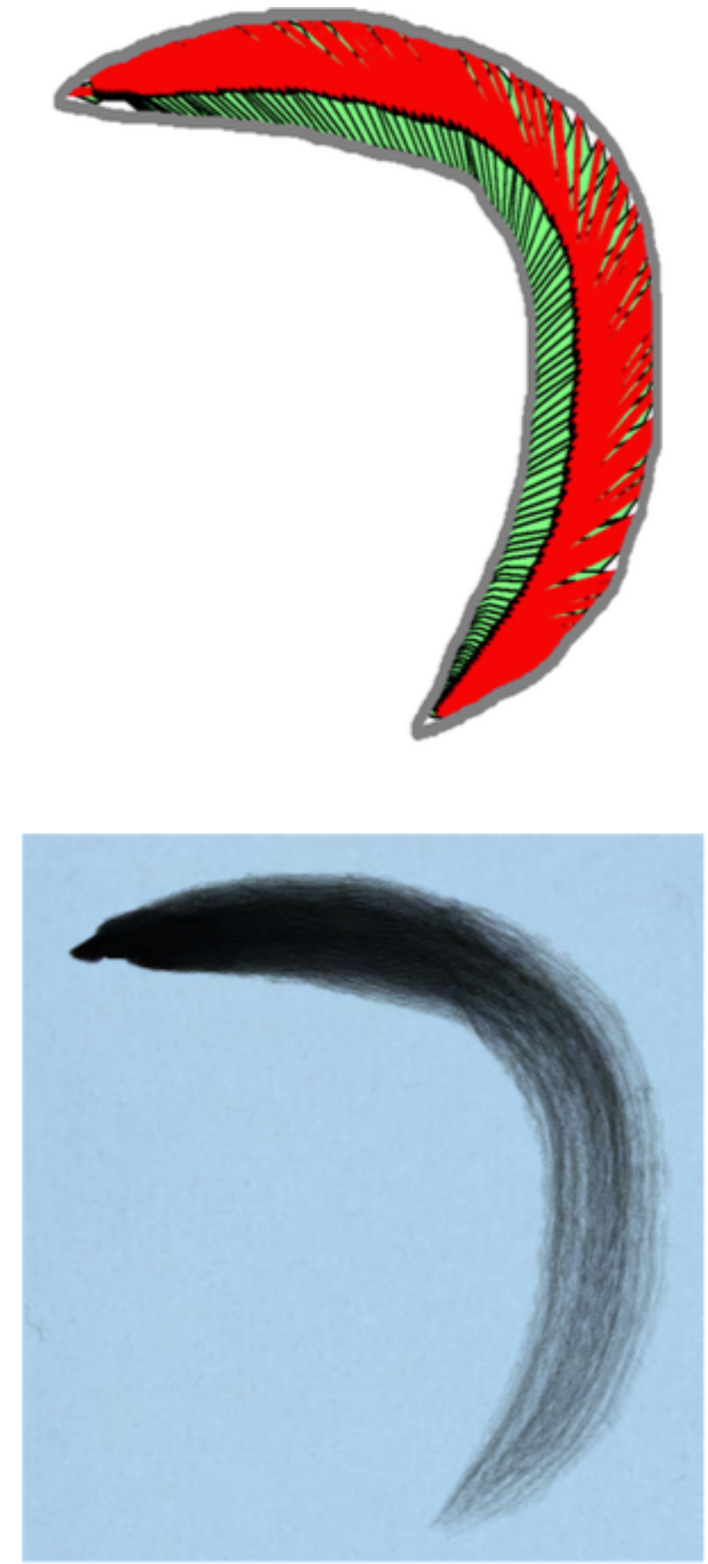
実験 1





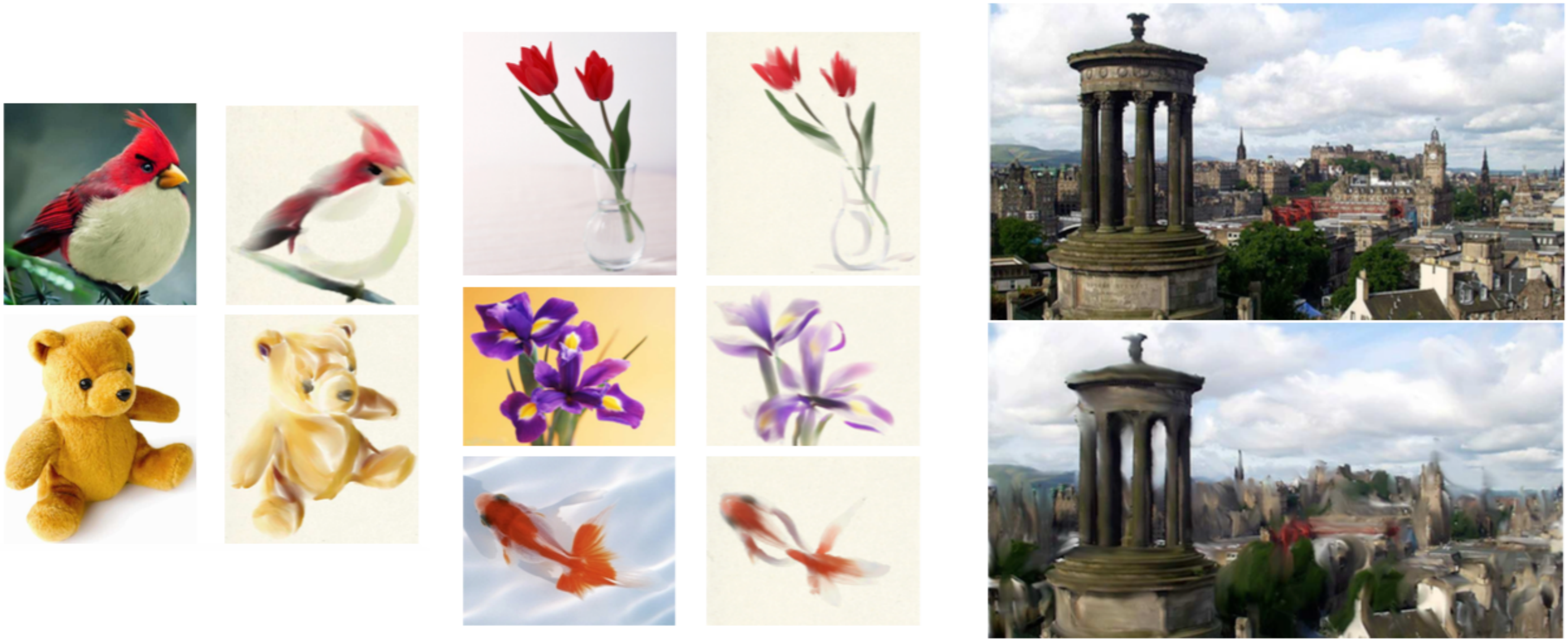
- 右図の上段は軌跡、下段は筆跡
- 学習を繰り返すと平均報酬が増加し塗りも良くなっている

#20



#43

実験2



輪郭は人間が手で描いたもの

本論文の貢献

- ✓ 墨絵に対して筆エージェントを設計して強化学習を適用した
- ✓ 筆をよく観察して状態・行動・報酬を丁寧に設計した
  - ✓ とくに周囲の状況（カーブの急峻さなど）に対して適切な状態空間を設計したのが重要な貢献
  - ✓ 基本図形を学習させ、エピソードの組み合わせで表現したことも貢献
- ✓ DPよりも自然な墨絵を描くことに成功した