

# 楕円による大まかな指定からの前景抽出

デンドログラム上の最適な切断と手がかりの信頼度に基づく重み付け

routersys

2026年10月

**概要** 被写体を楕円で大まかに指定したときに、画像からその前景のアルファマップを推定する方法を述べる。画像から単連結法でデンドログラムを作り、楕円に由来する事前分布、画素ごとの手がかり、境界項の和を最大にする切断を、動的計画法で厳密に求める。手がかりには、色、 $3\times 3$ パッチの最近傍、元の解像度の画像の局所的な平坦さを使う。平坦さは画素ごとの色変換で保たれるため、色転写を受けた合成画像でも前景と背景を分ける。さらに、楕円の内側と外側の分布のバタチャリヤ係数から手がかりの信頼度を正解なしで推定し、色の重みを調整する。動画編集ソフトウェアの出力から復元した前景とアルファマップを、CC0の背景に合成した同じキャラクターの髪型違いの立ち絵2種と、17体の立ち絵を16種の背景に合成した102場面（色転写0%と100%）で比較した。加えて、手法の設計に使っていない、公開素材の立ち絵17枚とCC0の背景16枚から作った102場面（色転写0%と100%）でも評価した。楕円をそのまま使うと前景のIoUは0.40から0.60であるのに対し、適応（色）の構成では0.68~0.93になる。テクスチャのみと色とパッチのみの優劣は画像データセットで異なり、両者を併用する固定重みは六つのデータセットのすべてで最良の構成とのIoUの差が0.012以内であった。信頼度に基づく色の重みの調整の効果は、IoUで最大0.004と小さい。画素単位のグラフカットは、17体の転写100%を除く五つのデータセットで木の切断よりIoUが0.005~0.024高く（信頼区間が0を含まないのは立ち絵2、公開素材の転写0%、公開素材の転写100%）、17体の転写100%では0.014低い（信頼区間は0を含まない）。同じスコアのもとでの最適化1回の計算時間は、木の動的計画法のほうが、Dinic法による最小カットの約44分の1、BoykovとKolmogorovのアルゴリズムの約1.4分の1である。

**キーワード** 前景抽出, 階層的領域分割, 単連結法, 動的計画法, 色転写, 被写界深度

---

## 1 はじめに

映像や画像の編集では、被写体を鮮明に保ったまま背景をぼかす処理がよく使われる。レンズの被写界深度による効果を模した処理であり、背景のぼかし量は画素ごとの合焦の度合いで決まる[1,2]。この処理を後から画像に加えるには、どの画素が合焦領域、すなわち前景に属するかを知る必要がある。前景の形を得る方法には、利用者が輪郭を丁寧に指定する方法と、深度や学習済みの分割モデルから得る方法[2]がある。しかし編集の途中で輪郭の指定を求めるのは負担が大きい。また、学習済みのモデルを使う場合は、重みの配布や管理が必要になる。

本稿は、利用者が画面上の楕円を動かして、合焦させたい範囲を大まかに示す操作を想定する。楕円は被写体の輪郭に沿ったものではなく、位置と大きさを合わせたものにすぎない。本稿が想定する誤差では、楕円の半径は、被写体の外接矩形に内接する楕円の0.8~1.6倍になり、中心は直径の12%まで動く。こ

のとき、楕円の内側には背景が多く含まれ、外側には前景がはみ出す。入力画像と楕円だけである。学習済みのモデルや追加の注釈を使わずに、前景のアルファマップを推定する問題を考える。

インタラクティブな前景抽出では、画素を頂点とするグラフ上のエネルギーを最小カットで最小化する方法が広く使われている[3,4]。この方法は、利用者が与えるシード（前景と背景の画素）から色の分布を推定し、画素ごとのデータ項と隣接画素間の平滑化項の和を最小にする。一方、画像から階層的な領域分割を作り、その木の上で領域の集合を選ぶ方法もある[5–7]。木の上の最適な切断は動的計画法で厳密に求め、更新の計算も軽い。ただし、木の節点が表す領域の境界の強さを目的関数にどう入れるかは自明でない。

本稿は、楕円からの前景抽出について、画素ごとの手がかり、境界項、最適化の三つを分けて設計し、それぞれの効果を立ち絵と背景の合成画像で測る。貢献は次の四つである。

1. 元の解像度の画像の局所的な平坦さを、前景と背景を分ける手がかりとして使う。画素ごとの色変換は、同じ値の隣接画素を同じ値のままにするため、平坦な領域は色転写を受けた後も平坦である。色とパッチの手がかりの識別性能が下がる合成画像でも、この手がかりは使える。
2. 楕円の内側と外側の分布のバタチャリヤ係数から、手がかりの信頼度を正解も学習も使わずに画像ごとに推定し、色の重みを調整する方法を試す。係数と手がかりの識別性能の関係を実測し、重みの調整の効果が小さいことを示す。
3. 領域の木の上で、領域ごとのスコアと境界のコストの和を動的計画法で厳密に最適化する枠組み[6,8,9]を、楕円からの前景抽出に適用する。単連結法のデンドログラムの上で、楕円に由来する事前分布、画素ごとの手がかり、境界項の和を最大にする切断を求める。境界項は、節点の領域の境界辺の類似度の平均で与え、すべての節点の値を最近共通祖先を使う集計により  $O(N \log N)$  で求める。境界項の三つの形と画素単位のグラフカットを、同じスコアのもとで比較する。
4. 動画編集ソフトウェアの出力から正解のアルファマップを復元する手順を示し、これと、17体のキャラクターの立ち絵（全身像の画像）を16種の背景に合成した102場面で手法を評価する。設計に使っていない立ち絵2と、公開されている素材だけで作ったデータセットの評価を、別に報告する。

以下、2章で関連研究を述べ、3章で問題と評価の設定を定める。4章で手法を、5章で実験を述べる。6章で結果の限界を議論し、7章でまとめる。

## 2 関連研究

### 2.1 前景抽出とグラフカット

Boykov と Jolly は、利用者が前景と背景の画素に印を付け、領域の情報と境界の情報を合わせたエネルギーを最小カットで大域的に最小化する方法を示した[3]。バイナリ変数のエネルギーのうち、グラフカットで最小化できる条件はKolmogorov と Zabih が与えている[10]。Rother らのGrabCut は、利用者が矩形で囲んだ領域から色分布を推定し、最小カットと分布の推定を繰り返して前景を求める[4]。Li らのLazy Snapping は、画像をあらかじめ過分割し、その領域の上でグラフカットを解くことで、利用者の操作に即座に応答する[11]。Kohli と Torr は、辺の重みが少し変わったときに、直前の最大流の解から次の解を速く求める動的グラフカットを示した[12]。Lempitsky らは、利用者の矩形が分割結果を十分に密に囲むという位相的な制約を、エネルギーの最小化に加える方法を示した[13]。

これらは、矩形が前景の外接矩形に近いことを前提にする。矩形が前景を緩く囲む場合は、YuらのLooseCutが、全体の類似度に関する制約と、外観が似た画素に同じラベルを促す項を加えたマルコフ確率場で扱う[14]。XuらのDeep GrabCutは、矩形を距離画像に変換して画像とともに畳み込みニューラルネットワークに入力し、大きすぎる矩形や小さすぎる矩形からも前景を求める。曲線による入力にも、再学習なしで拡張できる[15]。本稿の楕円は、被写体の外接矩形に内接する楕円に対し、縦横の半径を0.8~1.6倍に変え、中心を動かしたものである。前景が楕円からはみ出し、内側に背景が広く含まれるため、矩形が前景に接することに基づく制約は使えない。粗い入力から前景を求める問題設定は、これらの先行研究がすでに扱っている。本稿は、学習を使わず、デンドログラムの上で厳密に最適化する点で異なる。5章では、画素単位のグラフカットを比較手法として使う。

## 2.2 階層的な領域分割と木の上の切断

単連結法による階層的クラスタリングの結果は、最小全域木から得られる[16]。画像の隣接画素を辺とするグラフにこれを適用すると、辺の重みの小さい順に領域を併合するデンドログラムが得られる。重みが閾値以下の辺だけで連結した成分はquasi-flat zoneと呼ばれ、閾値を変えたときの階層は、最小全域木やsaliency mapと対応づけて調べられている[17]。Najmanらは、Kruskal法を変形して、このような階層を表す木を準線形の時間で作るアルゴリズムを与えた[18]。Soilleは、画素間の経路上の濃淡差に制約を課して階層的な分割を定めるconstrained connectivityを提案した[19]。FelzenszwalbとHuttenlocherは、領域の内部の変化と領域間の境界の強さを比べる併合の基準を持つグラフベースの分割法を示した[20]。SalembierとGarridoは、領域を併合した二分木(binary partition tree)を、画像処理や分割、情報検索のための表現として提案した[5]。Arbeláezらは、輪郭検出の結果から領域の階層を作る方法を示した[7]。

このような階層の上で、互いに重ならない領域を選ぶ研究がある。Guiguesらは、適合度の項と正則化の項からなるエネルギーを最小にする分割を、尺度の階層(scale-sets)から求めた[6]。Ravi KiranとSerraは、階層の上で最適な分割を選ぶ問題を階層的切断(hierarchical cuts)として整理し、エネルギーが一定の条件を満たすときに、線形時間の動的計画法で大域最適が得られることを示した[8]。HuらのTree-Cutは、領域の木の上で切断の確率分布を定め、動的計画法で厳密に推論する[21]。LempitskyらのPylonモデルは、領域の木の上で、領域ごとの単項のコストと境界のコストの和を最小にする、互いに重ならない領域の組を、グラフカットで大域最適に求める[9]。

本稿はこの系統に属する。木の上の最適な切断を動的計画法で厳密に解く点は、Guiguesら、Ravi KiranとSerra、Huらと共通する。領域ごとのスコアと境界のコストの和を、領域の木の上で大域最適に最小化する定式化は、Pylonモデルと共通する。違いは三つある。第一に、本稿の入力は楕円であり、楕円に由来する事前分布と画像の手がかりをスコアに入れる。第二に、境界のコストを節点ごとの加法的な項で与えるため、グラフカットを使わず、動的計画法だけで解ける。第三に、木の各節点の領域について、境界辺の類似度の平均を、最近共通祖先を使う集計で、すべての節点について求める。この集計と、4.6節の元の解像度の画像の平坦さの手がかりについて、本稿が調べた範囲では、同じ方法の先行研究を確認できていない。

## 2.3 アルファマットと正解の取得

アルファチャンネルによる画像の合成はPorterとDuffが定式化した[22]。自然画像のマッティングでは、Levinらが前景と背景の色の局所的な滑らかさを仮定し、疎な線形方程式を解いてアルファを求める方法を

示した[23]。マッティング手法の評価には、正解の付いた画像と、知覚に基づく誤差尺度を備えた評価基盤が使われている[24]。SmithとBlinnは、異なる二つの背景の前で同じ前景を撮影し、背景が全画素で異なれば、前景の色とアルファを復元できることを示した（三角測量マッティング）[25]。本稿の正解の復元は、背景を黒にした出力と、既知の背景を使った出力から、この考え方でアルファを求める（5.2節）。

## 2.4 色転写

Reinhardらは、一つの画像の色の平均と標準偏差を、別の画像のそれに合わせる色転写を提案した。計算はRudermanらの $\text{la}\beta$ 色空間で行い、チャンネルごとに独立に行う[26,27]。この変換は画素ごとに働くため、前景と背景の絵柄が異なる合成画像でも、前景の色を背景の色に近づけられる。本稿では、この方法を自作の実装で用いて、色転写の強さを変えた合成を評価に使い、色転写が色の手がかりに与える影響を調べる。

## 2.5 手がかりの信頼度と分布の離れ

ラベルが一部の標本にしかない場合や、ラベルに誤りが含まれる場合の分類は、正例とラベルなしのデータからの学習[28]や、ラベル雑音のもとでの学習として研究されている。Menonらは、二値ラベルが一定の割合で入れ替わる条件のもとで、誤ったラベルに対するAUCが、正しいラベルに対するAUCの一次関数になることを示した[29]。この結果は、観測したクラスごとの分布を真の二つの分布の混合として表す、Scottらの定式化[30]に基づく。Menonらは、この枠組みを *mutually contaminated distributions* と呼ぶ[29]。本稿は、楕円の中心側と外側の画素を、誤りを含みうる前景と背景の標本とみなす。

二つの分布の離れを測る量にバタチャリヤ係数 (Bhattacharyya coefficient) [31]があり、2クラスの分類のベイズ誤り率の上界を与える[32]。追跡では、Collinsらが、物体と背景の標本から対数尤度比の特徴を作り、二つの分布を分離する度合いで特徴の順位を決めて、よく分離する特徴を使った[33]。標本の分布から手がかりの識別力を正解なしで測る点は、本稿の信頼度と共通する。本稿は、分離の度合いにバタチャリヤ係数を使い、手がかりを選ぶのではなく、色の重みを連続的に調整する（4.7節）。

## 2.6 学習に基づく前景抽出

近年は、プロンプトを入力として任意の画像から領域を出力する大規模な学習済みモデルが示されている[34]。LiらのMatting Anythingは、このモデルの特徴を使い、箱、点、文章による指示からアルファマットを推定する[35]。このようなモデルは、学習データに由来する知識を使う点で、本稿の枠組みと前提が異なる。本稿は、学習を行わず、入力画像と楕円だけから前景を推定する手法の性質を調べるものであり、学習済みモデルとの比較は行っていない。

# 3 問題設定と評価

## 3.1 入力と出力

画像  $I: \Omega \rightarrow [0, 1]^3$  を入力とする。 $\Omega$  は  $W \times H$  画素の格子である。利用者の指定は楕円  $E = (c, r)$  で与えられる。 $c = (c_x, c_y)$  は中心、 $r = (r_x, r_y)$  は二つの半径である。画素  $p = (x, y)$  に対して、楕円に正規化した動径

$$\rho(p) = \sqrt{\left(\frac{x - c_x}{r_x}\right)^2 + \left(\frac{y - c_y}{r_y}\right)^2} \quad (1)$$

を定める。 $\rho < 1$  が楕円の内側である。出力は前景のアルファマップ  $\hat{\alpha}: \Omega \rightarrow [0, 1]$  である。

### 3.2 描画と評価指標

推定したアルファマップは、背景のぼかし描画に使う。前景の画素は  $I$  のまま残し、背景の画素だけを半径  $R$  の円板でぼかす。ぼかしに前景の色が混ざらないよう、正規化畳み込み (normalized convolution) [36]を使う。円板の核を  $G_R$ 、畳み込みを  $*$  で表すと、描画は

$$R(I, \alpha) = \alpha I + (1 - \alpha) \frac{G_R * ((1 - \alpha)I)}{\max(G_R * (1 - \alpha), \varepsilon)} \quad (2)$$

である。右辺の第2項は、背景の画素だけを重みとして平均した色になる。ここで  $\varepsilon = 10^{-4}$  とする。ぼかしによる光学的な効果は円板の点拡がり関数で近似しており、深度に応じた変化や、明るい光源の広がりには扱わない。

評価には二つの指標を使う。一つは前景の IoU (intersection over union) で、 $\hat{\alpha}$  と正解  $\alpha^*$  をそれぞれ 0.5 で二値化した集合の、共通部分と和集合の面積比である[37]。もう一つは、推定したアルファマップによる描画  $R(I, \hat{\alpha})$  と、正解による描画  $R(I, \alpha^*)$  の平均絶対誤差 (MAE) である。

$$\text{MAE} = \frac{255}{3|\Omega|} \sum_{p \in \Omega} \sum_{c=1}^3 \left| R(I, \hat{\alpha})_c(p) - R(I, \alpha^*)_c(p) \right| \quad (3)$$

評価は幅 640 画素の画像で行い、 $R = 12$  画素とする。MAE は 0 から 255 の値であり、IoU が前景の形の一致を測るのに対し、MAE は描画への視覚的な影響を測る。

### 3.3 楕円の誤差モデル

利用者が指定する楕円の誤差は、正解のアルファマップ  $\alpha^*$  から次のように模擬する。まず、 $\alpha^* > 0.5$  の外接矩形に内接する楕円  $E^* = (c^*, r^*)$  を作る。次に、各軸の倍率  $s_x, s_y$  と中心のずれ  $h_x, h_y$  を独立に一様乱数で引き、

$$c = c^* + 2(h_x r_x^*, h_y r_y^*), \quad r = (s_x r_x^*, s_y r_y^*) \quad (4)$$

とする。 $h$  は直径に対する割合である。主な誤差モデル M は、 $s \in [0.8, 1.6]$ 、 $h \in [-0.12, 0.12]$  とする。ほかに、倍率を対数一様に引く対称な誤差モデル S ( $s \in [0.74, 1.35]$ 、 $|h| \leq 0.12$ )、小さい誤差モデル T ( $s \in [0.9, 1.15]$ 、 $|h| \leq 0.05$ )、大きい誤差モデル L ( $s \in [0.7, 2.0]$ 、 $|h| \leq 0.20$ ) を用いる。誤差モデル M で、楕円の内側の面積のうち前景が占める割合は、立ち絵 1 と立ち絵 2 で平均 0.61、17 体で平均 0.42 である。外側にはみ出す前景の面積の割合は、立ち絵で平均 6%、17 体で平均 5% である。

## 4 提案手法

手法の全体を図 1 に示す。入力画像を縮小し、色の範囲を正規化した縮小画像から、単連結法でデンドログラムを作る。楕円に由来する事前分布と画素ごとの手がかりを足した値を画素のスコアとし、デンドロ

グラム上で、境界項を引いたスコアの和が最大になる切断を、動的計画法で求める。得られた低解像度のマスクは、評価の解像度のアルファマップに精緻化する。



図1 手法の概要。(a)は入力画像と楕円、(b)は単連結法のデンドログラムを高さで切った領域（領域の平均色で表示）、(c)は事前分布と手がかりの和（赤が前景側）、(d)はデンドログラム上の最適な切断、(e)は精緻化後の $\alpha$ 、(f)は $\alpha$ を使った背景のぼかし描画である。

#### 4.1 縮小画像

入力画像を幅  $L$  画素に縮小した画像を  $I_L$  とする（本稿では  $L = 256$ 、面積平均）。 $I_L$  の各色チャンネルを、2パーセントイルと98パーセントイルの間が0から1になるように線形に伸ばし、範囲の幅の下限を0.04、結果の上限を1.5として切り詰める。この正規化した画像  $w: \Omega_L \rightarrow \mathbb{R}^3$  を、以下では縮小画像と呼ぶ。画素数を  $N = |\Omega_L|$  で表す。

#### 4.2 単連結法のデンドログラム

縮小画像の4近傍の画素対を辺とし、辺  $e = (p, q)$  の重みを色の距離  $g_e = \|w(p) - w(q)\|_2$  とする。辺を重みの小さい順に並べ、素集合データ構造（union-find）で併合すると、葉が画素である二分木  $T$  が得られる。これは単連結法のデンドログラムであり、辺の整列と併合はKruskal法の変形として準線形の時間で行える[18]。内部節点  $v$  は二つの子  $l(v), r(v)$  を持ち、併合を起こした辺の重みを高さ  $h(v)$  として持つ。節点  $v$  は画素の集合  $R(v)$  を表し、その面積を  $A(v) = |R(v)|$  とする。単連結法の性質から、高さ  $t$  で木を切って得られる領域は、重みが  $t$  以下の辺だけで連結した成分に一致する[16]。この成分は、画像処理では quasi-flat zone と呼ばれ[17]、木はその階層にあたる。構築の計算量は、辺の整列で決まり  $O(N \log N)$  である。

#### 4.3 木の上の最適な切断

木の切断  $C$  を、互いに祖先と子孫の関係にない節点の集合と定める。切断が表す前景は  $F(C) = \bigcup_{v \in C} R(v)$  である。画素  $p$  のスコアを  $u(p)$ 、節点  $v$  の境界項を  $\pi(v) \geq 0$  とし、節点のスコアを

$$s(v) = \sum_{p \in R(v)} u(p) - \pi(v) \quad (5)$$

で定める。目的は、 $J(C) = \sum_{v \in C} s(v)$  を最大にする切断  $C$  を求めることである。 $u(p)$  が正の画素は前景に、負の画素は背景に寄与する。 $\pi(v)$  は、領域  $R(v)$  の境界のコントラストが低い（領域が周囲と似ている）ほど大きく、低コントラストの境界に沿った切断を抑える。

この最大化は、木の葉から根に向かう再帰で厳密に解ける。葉  $v$  では  $B(v) = \max \{0, s(v)\}$  とし、内部節点では

$$B(v) = \max \{ s(v), B(l(v)) + B(r(v)) \} \quad (6)$$

とする。 $B(v)$  は、 $v$  の部分木に含まれる切断のうち、スコアの和の最大値である。 $v$  の部分木の切断は、 $v$  自身を選ぶか、二つの子の部分木の切断を合わせたものであり、二つの子の部分木は互いに独立なので、再帰は最適性を保つ。根の値  $B(\text{root})$  が  $J$  の最大値である。各節点でどちらの選択が最大を与えたかを記録しておけば、根から辿って切断が復元できる。 $s(v)$  が与えられれば、計算量は  $O(N)$  である。図2に小さな例を示す。

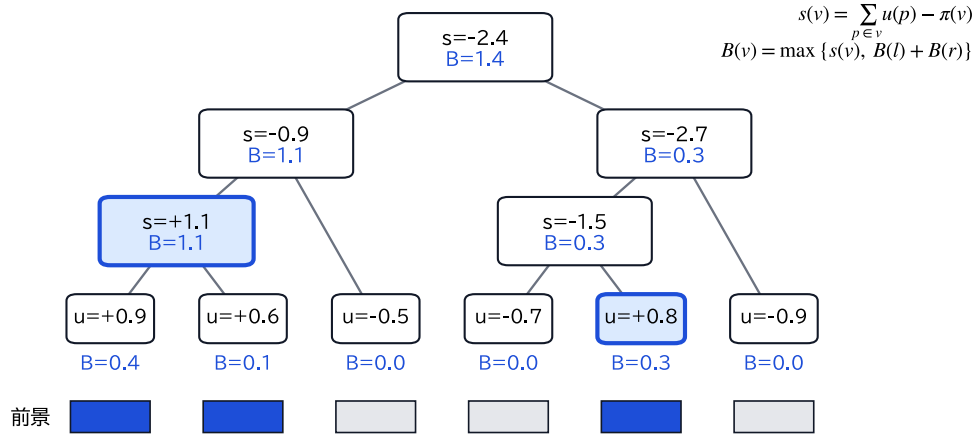


図2 木の上の最適な切断の例。葉の「 $u$ 」は画素のスコア、内部節点の「 $s$ 」は式(5)の節点のスコア、「 $B$ 」は式(6)で求めた最大値である。青い節点選ばれた切断で、下の帯は前景（青）と背景（灰）を示す。スコアが負の画素  $c$  を含む節点  $n2$  は、部分木の二つの選択の和よりスコアが小さいため選ばれない。

#### 4.4 楕円に由来する事前分布

楕円の内側ほど前景である可能性が高いとして、事前のスコアを定める。倍率  $a \in \{0.85, 1, 1.15\}$  の三つの楕円  $E_a = (c, ar)$  について、

$$\tilde{u}_a(p) = \text{clip} \left( \frac{1 - \rho(p) / a}{\kappa}, -1, 1 \right), \quad \tilde{u}_a(p) = \begin{cases} \gamma \tilde{u}_a(p) & \tilde{u}_a(p) > 0 \\ \tilde{u}_a(p) & \text{otherwise} \end{cases} \quad (7)$$

を作り、その平均  $u_E(p) = \frac{1}{3} \sum_a \tilde{u}_a(p)$  を事前のスコアとする。 $\kappa = 0.35$  は境界の幅、 $\gamma = 0.35$  は内側のスコアの縮小率である。誤差モデル  $M$  では楕円の内側にも背景が多く含まれるため、内側の正のスコアだけを弱め、外側の負のスコアは弱めない。三つの倍率の平均は、利用者の楕円の大きさが不確かであることに対応する。 $\gamma$  の影響は5.11節で調べる。

#### 4.5 境界項

辺  $e$  の類似度を  $\omega_e = \frac{1}{2} \sum_{\sigma \in \{0.05, 0.1\}} \exp(-g_e^2 / 2\sigma^2)$  とする。これは、グラフカットで隣接画素間の項に使われる重み[3]と同じ形である。重み  $g_e$  の小さい（色の近い）辺ほど  $\omega_e$  は1に近い。節点  $v$  の領域の境界辺の集合を  $\partial v = \{(p, q): p \in R(v), q \notin R(v)\}$  とし、境界辺の類似度の平均

$$\bar{\omega}(v) = \frac{1}{|\partial v|} \sum_{e \in \partial v} \omega_e \quad (8)$$

を使って、境界項を

$$\pi(v) = m \lambda_0 \frac{L}{128} \cdot 4\sqrt{A(v)} \bar{\omega}(v) \quad (9)$$

と定める。 $4\sqrt{A(v)}$  は、面積  $A(v)$  の正方形の周の長さである。境界辺の本数ではなくこの値を使うことで、輪郭が入り組んだ領域（髪の毛など）が過大なペナルティを受けるのを避ける。 $m$  は境界項の強さを調整する倍率で、開発用のデータセット（5.3節）で  $m = 0.5$  と選ぶ（5.8節）。積  $m\lambda_0$  だけが結果に影響するため、 $\lambda_0 = 14/3$  は  $m$  に吸収される定数である。係数  $L/128$  は、解像度が上がるとスコアの和が面積に比例して  $L^2$  で増える一方、周の長さが  $L$  でしか増えないため、両者の比を保つ補正である。

すべての節点について  $\sum_{e \in \partial v} \omega_e$  を求めるには、辺の両端点の最近共通祖先（LCA）を使う。各節点  $x$  に値  $\text{own}(x)$  を持たせ、辺ごとに、二つの端点に  $\omega_e$  を加え、両端点の最近共通祖先から  $2\omega_e$  を引く。ある節点  $v$  の部分木の  $\text{own}$  の総和を取ると、辺  $e$  の寄与は次のとおりになる。両端点が  $R(v)$  の外にあれば0である。一方の端点だけが内にあれば、最近共通祖先は  $v$  の祖先なので  $\omega_e$  である。両端点が内にあれば、最近共通祖先が  $v$  の部分木に入るため  $2\omega_e - 2\omega_e = 0$  になる。したがって、総和は境界辺の類似度の和になる。

$$\sum_{e \in \partial v} \omega_e = \sum_{x \in \text{sub}(v)} \text{own}(x), \quad \text{own}(x) = \sum_{e \ni x} \omega_e - 2 \sum_{e: \text{LCA}(e) = x} \omega_e \quad (10)$$

$\omega_e$  を1に置き換えれば  $|\partial v|$  も同様に求まる。辺の数は  $O(N)$  であり、LCAを、各節点から  $2^k$  個上の祖先を表に持つ方法（binary lifting）で求めると、全体で  $O(N \log N)$  である。

比較のため、併合高さだけを使う形（併合高さ型）と、境界辺の類似度の総和をそのまま使う形（総和型）も定義する。併合高さ型は  $\pi_{\text{wl}}(v) = m \lambda_0 \frac{L}{128} 4\sqrt{A(v)} \exp(-h(\text{parent}(v))^2 / 2\sigma^2)$ （上線は  $\sigma$  についての平均）である。単連結法では、 $v$  の領域から外へ出る辺のうち重みが最小のものが、親の節点を作る併合の辺になる。そのため、この形は境界辺のうち最も類似度が高い（色の差が最小の）一本だけを見ている。総和型は  $\pi_{\Sigma}(v) = m \lambda_0 \sum_{e \in \partial v} \omega_e$  である。

#### 4.6 画素ごとの手がかり

楕円の内側の中心側  $K = \{\rho < 0.5\}$  と、外側  $B = \{\rho > 1.3\}$  の画素を、それぞれ前景と背景の標本とみなす。手がかり  $c$  ごとに、この二つの標本から特徴量  $x_c(p)$  の分布  $\hat{P}_K^c, \hat{P}_B^c$  を推定し、対数尤度比を  $[-1, 1]$  に正規化してスコアにする。

$$e_c(p) = \frac{1}{2} \text{clip} \left( \ln \frac{\hat{P}_K^c(x_c(p))}{\hat{P}_B^c(x_c(p))}, -2, 2 \right) \in [-1, 1] \quad (11)$$

中心側の標本が前景だけとは限らず、外側の標本も背景だけとは限らない。したがって  $e_c$  は、この二つの標本に対する識別性能にすぎない。

**色。** 縮小画像のRGBを各チャンネル8段階に量子化し、512段階のヒストグラムを両標本から作る。ヒストグラムは $3 \times 3 \times 3$ の近傍で和を取って滑らかにし、正規化してから  $10^{-5}$  を足して比を取る。

**パッチ。** 各画素について、 $3 \times 3$  近傍のRGB値を並べた27次元のベクトル  $P(p)$  を作る。中心側から最大3000画素、外側から最大で中心側の標本数の8倍（下限20）の画素を無作為に選び、それぞれ標本集合  $S_K$ 、 $S_B$  とする。画素  $p$  から各標本集合への最近傍距離

$$d_S(p) = \min_{q \in S, \|p-q\|_\infty > 2} \|P(p) - P(q)\|_2 \quad (12)$$

を求める。画素自身とその近傍を除くことで、標本の画素が自分自身を最近傍にするのを避ける。スコアは  $e_{\text{pat}}(p) = \frac{1}{2} \text{clip} \left( 4 \ln \frac{d_B(p) + 10^{-3}}{d_K(p) + 10^{-3}}, -2, 2 \right)$  である。画像から標本集合への距離を使った識別は、量子化を介さない最近傍による画像分類（NBNN）でも使われている[38]。最近傍法の漸近的な誤り率はベイズ誤り率の2倍以下である[39]。

**テクスチャ。** 元の解像度の8ビット画像  $J$  について、各画素  $p$  とその  $3 \times 3$  近傍の画素  $q$  の色の差の最大値

$$m(p) = \max_{q \in N_{3 \times 3}(p)} \max_c |J_c(p) - J_c(q)| \in \{0, \dots, 255\} \quad (13)$$

を求める。 $m(p) \leq 2$  のとき1となる指示関数  $f(p) = 1[m(p) \leq 2]$  と、 $\min(m(p), 32) / 32$  の二つの特徴を作り、縮小画像の格子に面積平均で縮小して、さらに  $3 \times 3$  の平均で滑らかにしたものを  $F(p)$ 、 $M(p)$  とする。 $(F, M)$  を各6段階に量子化した36段階の同時ヒストグラムを両標本から作り、式(11)で対数尤度比を求める（頻度に0.5を足す）。

この手がかりが、色転写の影響を受けにくい理由は次のとおりである。色転写は、画素の色だけから出力の色を決める変換  $\phi$  ( $J'(p) = \phi(J(p))$ ) として表される。このとき  $J(p) = J(q)$  ならば  $J'(p) = J'(q)$  であり、したがって  $m(p) = 0$  の画素は、変換の後も  $m'(p) = 0$  になる。色の差が完全に0の領域は、どのような画素ごとの色変換を受けても保たれる。一定色で塗られた領域をもつ前景と、絵画的な濃淡やノイズをもつ背景の組では、この平坦さが前景と背景を分ける。合成された領域の間でノイズの性質が異なることを手がかりにする点は、画像の改ざん検出で用いられるノイズの不整合の解析[40]と共通する。縮小した画像では平坦な領域が平均されて失われるため、この手がかりは元の解像度の画像から計算する。図3に例を示す。

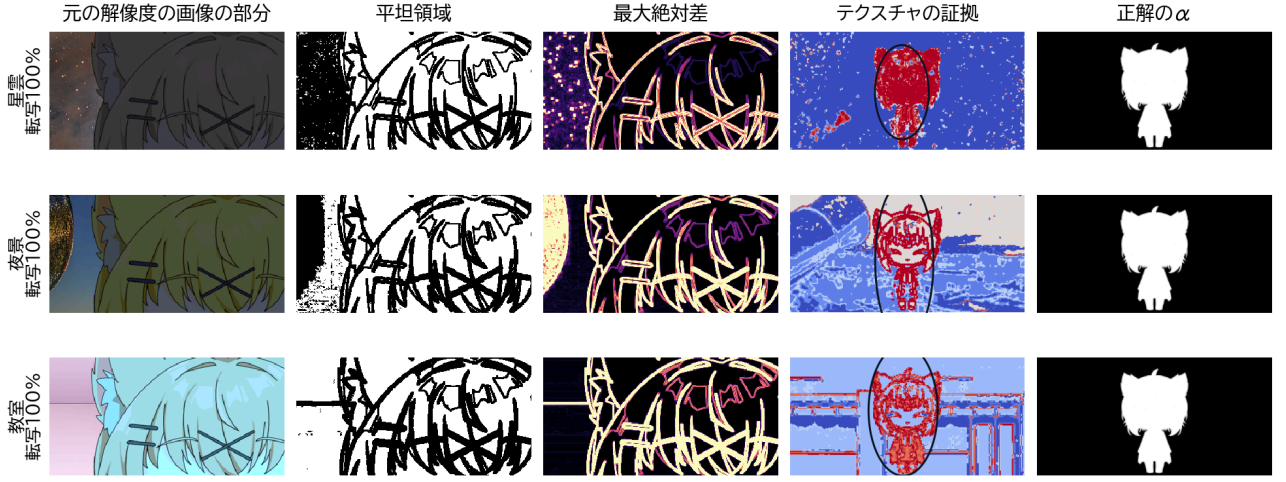


図3 テクスチャの手がかりの可視化（立ち絵1、転写100%）。左から、元の解像度の画像の髪と背景の境界付近、平坦領域（白が  $m(p) \leq 2$  の画素）、 $3 \times 3$  近傍内の最大絶対差（32で打ち切り）、テクスチャのスコア  $e_{\text{tex}}$ （赤が前景側、黒い楕円は指定した楕円の一例）、正解の  $\alpha$  である。

#### 4.7 手がかりの併用と信頼度に基づく重み

画素のスコアは、事前のスコアと手がかりの重み付き和とする。

$$u(p) = u_E(p) + \omega_{\text{tex}} e_{\text{tex}}(p) + \omega_{\text{col}} e_{\text{col}}(p) + \omega_{\text{pat}} e_{\text{pat}}(p) \quad (14)$$

重みを固定値  $(\omega_{\text{tex}}, \omega_{\text{col}}, \omega_{\text{pat}}) = (1.0, 0.25, 0.3)$  にしたものを、固定重みと呼ぶ。5章に示すとおり、色とパッチを含む重みと、テクスチャだけの重みのどちらが優れるかは、データセットで異なる。色転写で色の手がかりの識別性能が下がる場合があるためである。そこで、手がかりの信頼度を正解を使わずに画像ごとに推定し、色の重みを調整する。

手がかり  $c$  の特徴ベクトル（色は3次元、テクスチャは2次元の  $(F, M)$ ）に対して、 $K$  と  $B$  の標本にそれぞれ正規分布  $N(\mu_K, \Sigma_K)$ 、 $N(\mu_B, \Sigma_B)$  を当てはめ、バタチャリヤ係数[31]を閉じた形で求める。

$$\text{BC}_c = \exp(-D_B), \quad D_B = \frac{1}{8}(\mu_K - \mu_B)^\top \Sigma^{-1}(\mu_K - \mu_B) + \frac{1}{2} \ln \frac{\det \Sigma}{\sqrt{\det \Sigma_K \det \Sigma_B}}, \quad \Sigma = \frac{\Sigma_K + \Sigma_B}{2} \quad (15)$$

$\text{BC}_c$  は0から1の値で、小さいほど二つの標本の分布が離れている。事前確率を  $\pi_1, \pi_2$  とすると、二つのクラスの分類のベイズ誤り率は  $\sqrt{\pi_1 \pi_2} \text{BC}$  以下である[32]。係数は手がかりの種類の間で直接比べられない。そこで、色転写の影響を受けにくいテクスチャの係数を基準にして、色の重みを

$$\omega_{\text{col}} = 0.5 \max \left( 0, 1 - \frac{\text{BC}_{\text{col}}}{\text{BC}_{\text{tex}}} \right) \quad (16)$$

と定める。色の分布の離れがテクスチャの分布の離れより大きいほど重みは大きく、同程度以下なら0になる。テクスチャの重みは1.0、パッチの重みは0.3に固定する。パッチの係数は、5.7節に示すとおり識別性能との対応が弱いので、重みの調整に使わない。パッチの重みも同じ式で調整した場合は、5.4節で別に報告する。

## 4.8 精緻化

切断から得た幅  $L$  の二値マスク  $m_L$  は、評価の解像度のアルファマットに引き上げ、縁をなめらかにする精緻化を行う。まず、画素  $p$  の色  $I(p)$  と、 $p$  に近い低解像度の4画素  $q$  の色  $I_L(q)$  の距離を使って

$$\alpha_0(p) = \frac{\sum_q \varphi(p, q) \exp \left( -\|I(p) - I_L(q)\|^2 / 2\sigma_r^2 \right) m_L(q)}{\sum_q \varphi(p, q) \exp \left( -\|I(p) - I_L(q)\|^2 / 2\sigma_r^2 \right)} \quad (17)$$

を計算する（ジョイントバイラテラルアップサンプリング、joint bilateral upsampling[41]）。 $\varphi$  は双一次補間の重み、 $\sigma_r = 0.12$  である。次に、 $0.02 < \alpha_0 < 0.98$  の画素とその2画素外側の帯の中だけで、 $\alpha_0$  のロジットに色の対数尤度比を加えて補正する。補正後の値は  $\alpha_1 = \text{sigmoid} \left( 6(\alpha_0 - \frac{1}{2}) + 3\tilde{\ell} \right)$  で、 $\tilde{\ell} \in [-1, 1]$  は、マスクの内側と外側の画素の色から作った、チャンネルごと32段階のヒストグラムによる対数尤度比（絶対値6で打ち切り、6で割ったもの）である。帯の外では  $\alpha_0$  をそのまま使う。最後に、画像  $I$  をガイド画像とする色つきのガイドドフィルタ（guided filter）[42]（半径2、 $\epsilon = 10^{-3}$ ）をかけて  $\hat{\alpha}$  とする。

## 4.9 計算量

デンドログラムの構築と境界項の計算は  $O(N \log N)$ 、動的計画法は  $O(N)$  である。計算量が最も大きいのはバッチの最近傍探索で、標本の数  $n_K, n_B$  として  $O(N(n_K + n_B))$  である。3000画素と最大24000画素の標本を使う設定では、 $L = 256$ （ $N$  は約3.7万）で、全画素への距離は約  $10^9$  回の27次元の距離計算になる。色とテクスチャの手がかりはヒストグラムの数え上げだけで、 $O(N)$  である。計算は、楕円が変わっても作り直す必要のない部分と、楕円の変更のたびに計算する部分に分かれる。デンドログラムと境界項は画像だけで決まるため、前者にあたる。楕円を動かす操作では、 $K, B$ 、事前のスコアと手がかりが変わるので、これらを求め直し、そのあとに動的計画法を  $O(N)$  で再計算する。

---

### アルゴリズム1 楕円からの前景抽出

---

1. 入力画像から縮小画像を作り（4.1節）、4近傍の辺で単連結法のデンドログラムを構築する（4.2節）。
  2. すべての節点について  $A(v)$  と式(8)の  $\omega(v)$  をLCAを使って求め、境界項  $\pi(v)$  を得る（4.5節）。
  3. 楕円から  $K, B$  と事前のスコア  $u_E$  を作る。三つの手がかりのスコア  $e_c$  を求め、バタチャリヤ係数から色の重みを定めて、 $u$  を得る（4.4節、4.6節、4.7節）。
  4. 葉から根へ  $s(v)$  と  $B(v)$  を求め、根から切断を復元して低解像度のマスクを得る（4.3節）。
  5. マスクを精緻化して  $\hat{\alpha}$  を得て（4.8節）、ぼかし描画に使う（3.2節）。
- 

## 5 実験

### 5.1 評価に用いる画像

六つの画像データセットを使う（表1）。立ち絵1、立ち絵2、17体、公開素材の四つの画像群があり、17体と公開素材は、色転写の量が0%と100%のデータセットに分ける。

**立ち絵1。**等身の小さい猫耳のキャラクターで、輪郭線と一様な塗りからなる。動画編集ソフトウェアYMM4で書き出した画像から、前景の色とアルファマットを復元し（5.2節）、CC0の背景に合成した1280×640画素の画像である。前景の色を背景の色に近づける処理を、色転写（color transfer）と呼ぶ[26]。

星雲の写真の背景について、色転写の強さを0、25、50、75、100%に変えた5条件と、強さを100%にして背景を7種に替えた7条件を使う。計12条件である。

**立ち絵2。**立ち絵1と同じキャラクターで、髪型だけが異なる画像を同じ手順で書き出し、星雲の写真を含む8種のCC0の背景に、強さ100%の色転写をかけて合成した8条件である。背景は立ち絵1と一部が異なる。このデータセットは手法の設計と定数の選択に使っていないが、キャラクターと絵柄は立ち絵1と共通である。

**17体。**17体の立ち絵（アルファチャンネル付きの全身像）を、16種の背景画像から無作為に選んだ背景に、画面の高さの55～92%の大きさで合成した102場面である。背景は、オペレーティングシステムに付属する壁紙8枚と、イラスト調の背景8枚である。画像は1280×720画素で、合成後に8ビットに量子化する。色転写は自作の実装で、Reinhardらの方法[26]に従い、前景の画素の色の平均と標準偏差を、 $\alpha\beta$ 色空間[27]で背景全体のそれに合わせる。強さ0（変換なし）と強さ1（完全に合わせる）の2通りを作り、それぞれ独立したデータセットとして扱う。

**公開素材。**公開されている立ち絵素材から選んだ17枚を、17体と同じ手順で合成した102場面である。素材は、Breezy氏の「Visual Novel Sprites」（CC0）の4人（うち1人は衣装が2種類）の5枚と、4cher氏（Cyanca）の「[FREE-TO-USE] Visual Novel Sprites」（セット8を除きCC BY 4.0）の7セットの12枚である。どれもアルファチャンネル付きで、画像の下端で切れた立ち絵であり、絵柄は手描きのアニメ調から陰影の強い塗りまで異なる。各セットの表情の差分のうち、無表情に近いものを1枚ずつ選んだ。立ち絵は画像の下端に接する位置に置き、大きさ、横位置、色転写は17体と同じ方法で決める。背景は、CC0で公開されている16枚（写真8枚、絵画4枚、アニメ調の背景4枚）から無作為に選ぶ。このデータセットは手法の設計と定数の選択に使っておらず、立ち絵も背景も誰でも入手できるため、評価を再現できる。出典は謝辞に示す。

**表1** 評価に用いる画像データセット。試行は、誤差モデルMで引いた楕円の一つを指す。色転写の強さは、0%が変換なし、100%が前景の色の平均と標準偏差を背景に完全に合わせる場合である。

| データセット       | 画像の大きさ   | 条件・場面 | 試行数     | 正解の取得     | 開発用 |
|--------------|----------|-------|---------|-----------|-----|
| 立ち絵1         | 1280×640 | 12条件  | 96（各8）  | 出力から復元    | はい  |
| 立ち絵2         | 1280×640 | 8条件   | 80（各10） | 出力から復元    | いいえ |
| 17体・色転写0%    | 1280×720 | 102場面 | 204（各2） | アルファチャンネル | はい  |
| 17体・色転写100%  | 1280×720 | 102場面 | 204（各2） | アルファチャンネル | はい  |
| 公開素材・色転写0%   | 1280×720 | 102場面 | 204（各2） | アルファチャンネル | いいえ |
| 公開素材・色転写100% | 1280×720 | 102場面 | 204（各2） | アルファチャンネル | いいえ |

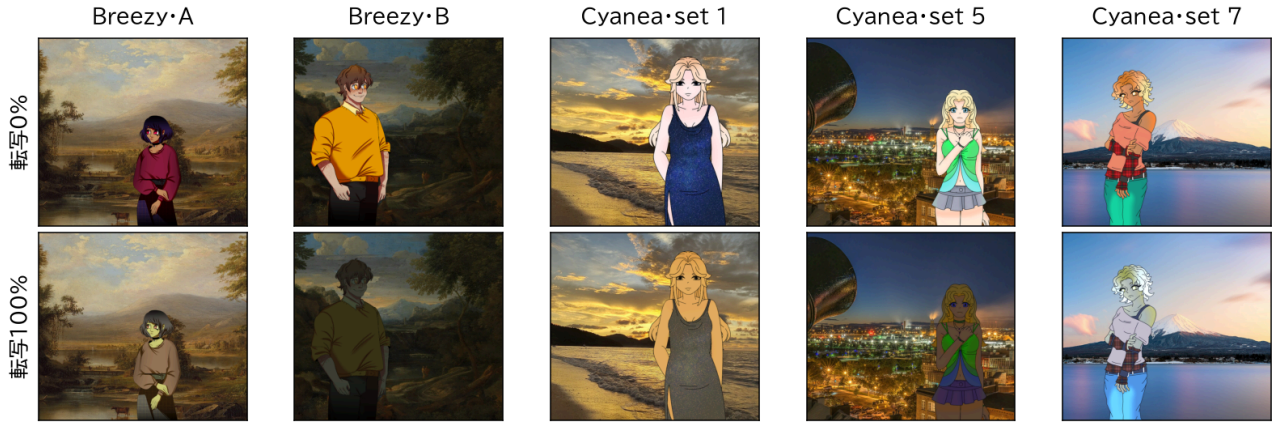


図4 公開素材の合成例。上段は色転写0%、下段は色転写100%である。立ち絵と背景の出典は謝辞に示す。

## 5.2 正解のアルファマットの復元

YMM4が書き出す画像は透過を保たない。そこで、同じ編集状態から三種類の画像を書き出し、正解のアルファマットを復元する。すべての背景を含む通常出力  $C$ 、背景のレイヤーを非表示にして書き出した出力  $K$ 、立ち絵のレイヤーを非表示にして書き出した出力  $B$  である。立ち絵の色を  $F$ 、アルファを  $\alpha$  とすると、 $C = \alpha F + (1 - \alpha)B$ 、 $K = \alpha F$ （背景が黒になる）であり、 $C - K = (1 - \alpha)B$  が各画素で成り立つ。色チャンネルについての最小二乗解は

$$\alpha = 1 - \frac{\sum_c B_c (C_c - K_c)}{\sum_c B_c^2} \quad (18)$$

である。これは、黒と既知の背景の二つの背景の前で同じ前景を撮影して、前景のアルファと色を復元する三角測量マッピング[25]の式である。 $\alpha$  を0から1に切り詰め、0.01未満を0にする。

得られた  $\alpha$  から  $K + (1 - \alpha)B$  を再構成すると、立ち絵1の例で、 $C$  との差は平均0.001、最大2.0（0から255の値）であった。画面の端にある黒い柱の近傍に出る誤差は、キャラクターの列の範囲の外を0にして除いた。立ち絵の位置と大きさは条件によらず同じであり、 $\alpha$  は条件の間で共通に使う。前景の色は、元の色の出力から  $K$ （背景が黒）を取り、 $\alpha$  で割って求める。立ち絵2の元の色の出力は、拡大率と位置が  $\alpha$  の書き出しと異なるため、輪郭が一致するように拡大率と位置を合わせてから使う。 $\alpha$  が0.5未満の画素の色は、近い画素の色で補う。これらを、CC0の背景に、17体と同じ式で合成する（図5）。



図5 YMM4の出力からの前景の復元と合成（立ち絵1）。(a)は背景を非表示にした出力  $K$ 、(b)は式(18)で復元した  $\alpha$ 、(c)は  $K$  を  $\alpha$  で割って求めた前景の色  $F$ 、(d)と(e)は星雲の写真の背景に合成した画像（色転写の強さが0%と100%）である。

## 5.3 比較する構成と手順

同じデンドログラム、境界項、精緻化を使い、画素のスコアだけを変えた構成を比較する（表2）。

**表2** 比較する構成。画素のスコアは式(14)の  $u(p) = u_E + \omega_{\text{tex}}e_{\text{tex}} + \omega_{\text{col}}e_{\text{col}} + \omega_{\text{pat}}e_{\text{pat}}$  で、重みは  $(\omega_{\text{tex}}, \omega_{\text{col}}, \omega_{\text{pat}})$  の順に示す。

| 構成         | 重み                                                    | 内容                                         |
|------------|-------------------------------------------------------|--------------------------------------------|
| 楕円をそのまま使用  | —                                                     | 楕円の内側を前景とした二値マスクに、精緻化をかける                  |
| 事前分布のみ     | (0, 0, 0)                                             | 楕円に由来する事前分布 $u_E$ だけをスコアにする                |
| テクスチャのみ    | (1.0, 0, 0)                                           | 元の解像度のテクスチャの手がかりだけを加える                     |
| 色とパッチのみ    | (0, 0.5, 0.6)                                         | 色とパッチの手がかりだけを加える                           |
| 固定重み       | (1.0, 0.25, 0.3)                                      | 三つの手がかりを固定の重みで併用する                         |
| 適応 (色)     | (1.0, $\omega_{\text{col}}$ , 0.3)                    | 色の重みを式(16)で画像ごとに定める                        |
| 適応 (色とパッチ) | (1.0, $\omega_{\text{col}}$ , $\omega_{\text{pat}}$ ) | パッチの重みも同じ式で、基準を0.6として定める                   |
| 正解スコア      | —                                                     | $u = 2\alpha^* - 1$ をスコアにする。木の表現と精緻化の限界を示す |

3章の誤差モデルMで楕円を引き、各画像について表の試行数だけ繰り返す。各構成は、同じ試行に対して、同じ手がかりのスコアを使って評価する。差の信頼区間は、場面（立ち絵1と立ち絵2では条件）を単位として復元抽出を2000回行うブートストラップ法で求める。立ち絵1と17体を、手法の設計と定数の選択に使う開発用のデータセットと呼ぶ。最適な定数は開発用のデータセットで選び、立ち絵2と公開素材は選択に使わない。定数の一覧は付録Aに示す。

精度の評価はPythonとNumPyによる参照実装で行い、木の動的計画法とグラフカットの計算時間だけは、同じ入力に対するC#の実装で測った。

## 5.4 主な結果

表3と表4に、誤差モデルMでの結果を示す。図6に、六つのデータセットのIoUと、場面を単位とするブートストラップ法による95%信頼区間を示す。

**事前分布と手がかりの効果。**楕円をそのまま使うと、IoUは立ち絵1で0.575、立ち絵2で0.588、17体の転写0%と100%で0.406と0.404である。デンドログラムと境界項だけを使う事前分布のみの構成は、これを+0.184、+0.101、+0.286、+0.161だけ上げる。四つのデータセットのすべてで、信頼区間は0を含まない。テクスチャの手がかりを加えると、事前分布のみに対してさらに+0.154、+0.193、+0.058、+0.086上がる。

**手がかりの順位はデータセットで異なる。**立ち絵1の全12条件では、テクスチャのみのIoUが0.913で、色とパッチのみの0.867を上回る。色とパッチのみは色転写の強さが増えるほど悪化し、星雲の背景では、転写0%の0.960から100%の0.815に下がる。テクスチャのみは0.965から0.961にとどまる。17体の転写0%では、色とパッチのみが0.791で、テクスチャのみの0.750を上回り、順位が立ち絵1と逆になる。転写100%では、テクスチャのみが0.651、色とパッチのみが0.630で、固定重みが0.677で最良に近い。17体の転写100%では、色の手がかりの識別性能が下がっても、パッチの手がかりはテクスチャより識別性能が高く（5.7節）、併用が有効だったと考えられる。

固定重みのテクスチャのみに対するIoUの差は、立ち絵1で+0.016 [+0.006, +0.026]、設計に使っていない立ち絵2で+0.004 [-0.003, +0.013]、17体の転写0%で+0.042 [+0.032, +0.054]、転写100%で+0.026

[+0.017, +0.037] である。信頼区間が0を含まないデータセットでは、いずれも固定重みが上回る。固定重みは、四つのデータセットのどれでも、最良の構成との差が0.012以内に収まる。

重みの適応の効果。適応（色）の固定重みに対するIoUの差は、立ち絵1で-0.003 [-0.006, +0.001]、立ち絵2で+0.001 [0.000, +0.003]、17体の転写0%で-0.004 [-0.007, -0.001]、転写100%で+0.004 [-0.001, +0.008] である。大きさは最大でも0.004にとどまり、符号はデータセットで異なる。パッチの重みも調整した適応（色とパッチ）は、17体で+0.006 [+0.001, +0.012] と+0.011 [+0.005, +0.018] と、信頼区間が0を含まない向上を示す。立ち絵1では+0.002 [-0.001, +0.005] で、信頼区間が0を含む。立ち絵2では+0.003 [0.000, +0.008] である。MAEでも、適応（色とパッチ）は17体で小さくなり（転写0%で-0.07 [-0.10, -0.05]）、ほかのデータセットでは差が小さい。

上限との差。正解スコアを用いたときのIoUは、立ち絵1で0.966、立ち絵2で0.944、17体の転写0%と100%で0.961と0.870である。これは、デンドログラムと精緻化が表現できる形の限界を示す。適応（色）との差は、それぞれ0.040、0.056、0.173、0.189であり、手がかりの精度に改善の余地が残る。特に17体では差が0.17以上あり、本稿の手がかりと楕円だけでは前景の形を十分に決められない。

手がかりが働かない場面。17体の転写100%には、背景がほぼ一様な壁紙で、強さ100%の色転写により前景が背景と同じ色の様な塗りになる場面が6場面（12試行、全204試行の6%）ある。この場面では、三つの手がかりの楕円内のAUCがすべて0.5で、バタチャリヤ係数が1になり（図8の右下の点）、手がかりを使う構成のIoUの平均は0.000である。これらの場面を除くと、17体の転写100%の固定重みのIoUは0.677から0.719になる。

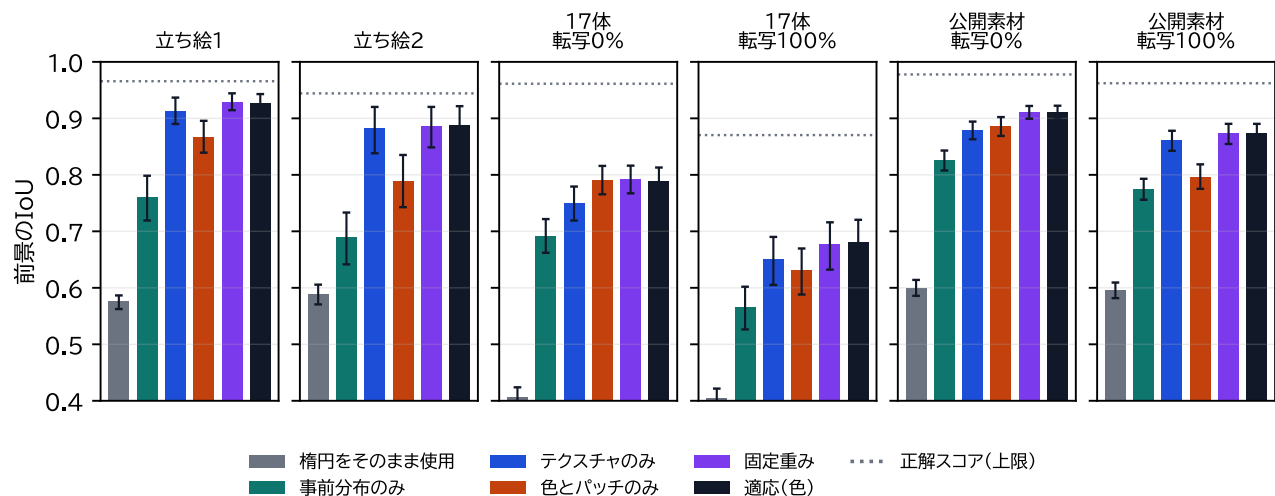


図6 データセットごとの前景のIoU（誤差モデルM、精緻化後）。棒は場面を単位とする平均、縦線は95%信頼区間（ブートストラップ法）、点線は正解スコアを用いた上限である。

表3 前景のIoU（各行の最大を太字で示す。正解スコアは上限として別に示し、太字の対象に含めない）。

| 条件              | 楕円を<br>そのまま | 事前分布<br>のみ | テクスチャ<br>のみ | 色と<br>パッチのみ | 固定<br>重み     | 適応<br>(色) | 適応 (色と<br>パッチ) | 正解<br>スコア |
|-----------------|-------------|------------|-------------|-------------|--------------|-----------|----------------|-----------|
| 立ち絵1・星雲・転写0%    | 0.580       | 0.874      | 0.965       | 0.960       | 0.964        | 0.964     | <b>0.965</b>   | 0.963     |
| 立ち絵1・星雲・転写25%   | 0.573       | 0.798      | 0.952       | 0.903       | <b>0.952</b> | 0.951     | 0.952          | 0.957     |
| 立ち絵1・星雲・転写50%   | 0.607       | 0.839      | 0.948       | 0.900       | <b>0.951</b> | 0.946     | 0.942          | 0.958     |
| 立ち絵1・星雲・転写75%   | 0.526       | 0.739      | 0.952       | 0.785       | 0.951        | 0.954     | <b>0.954</b>   | 0.965     |
| 立ち絵1・星雲・転写100%  | 0.575       | 0.700      | 0.961       | 0.815       | <b>0.961</b> | 0.958     | 0.957          | 0.979     |
| 立ち絵1・他の7背景・100% | 0.578       | 0.738      | 0.883       | 0.863       | 0.910        | 0.906     | <b>0.915</b>   | 0.967     |
| 立ち絵1・全12条件      | 0.575       | 0.760      | 0.913       | 0.867       | 0.929        | 0.926     | <b>0.931</b>   | 0.966     |
| 立ち絵2・8背景・100%   | 0.588       | 0.689      | 0.882       | 0.788       | 0.887        | 0.888     | <b>0.890</b>   | 0.944     |
| 17体・転写0%        | 0.406       | 0.692      | 0.750       | 0.791       | 0.792        | 0.789     | <b>0.798</b>   | 0.961     |
| 17体・転写100%      | 0.404       | 0.565      | 0.651       | 0.630       | 0.677        | 0.681     | <b>0.688</b>   | 0.870     |
| 公開素材・転写0%       | 0.600       | 0.825      | 0.879       | 0.885       | 0.911        | 0.912     | <b>0.917</b>   | 0.978     |
| 公開素材・転写100%     | 0.596       | 0.775      | 0.861       | 0.796       | 0.873        | 0.874     | <b>0.874</b>   | 0.962     |

表4 背景のぼかし描画のMAE（0～255。小さいほどよい。各行の最小を太字で示す）。

| 条件              | 楕円を<br>そのまま | 事前分布<br>のみ | テクスチャ<br>のみ | 色と<br>パッチのみ | 固定<br>重み    | 適応<br>(色) | 適応 (色と<br>パッチ) | 正解<br>スコア |
|-----------------|-------------|------------|-------------|-------------|-------------|-----------|----------------|-----------|
| 立ち絵1・星雲・転写0%    | 1.12        | 0.98       | 0.25        | 0.25        | 0.27        | 0.28      | <b>0.25</b>    | 0.21      |
| 立ち絵1・星雲・転写25%   | 1.02        | 0.88       | <b>0.26</b> | 0.46        | 0.28        | 0.29      | 0.27           | 0.18      |
| 立ち絵1・星雲・転写50%   | 0.73        | 0.54       | <b>0.26</b> | 0.38        | 0.28        | 0.26      | 0.26           | 0.17      |
| 立ち絵1・星雲・転写75%   | 0.88        | 0.51       | 0.22        | 0.46        | 0.23        | 0.18      | <b>0.17</b>    | 0.15      |
| 立ち絵1・星雲・転写100%  | 0.73        | 0.53       | 0.19        | 0.38        | <b>0.18</b> | 0.19      | 0.19           | 0.14      |
| 立ち絵1・他の7背景・100% | 1.63        | 1.32       | 0.67        | 0.75        | 0.56        | 0.57      | <b>0.52</b>    | 0.28      |
| 立ち絵1・全12条件      | 1.33        | 1.06       | 0.49        | 0.60        | 0.43        | 0.43      | <b>0.40</b>    | 0.24      |
| 立ち絵2・8背景・100%   | 1.67        | 1.36       | 0.81        | 0.95        | 0.76        | 0.75      | <b>0.73</b>    | 0.30      |
| 17体・転写0%        | 1.87        | 1.28       | 1.14        | <b>0.78</b> | 0.90        | 0.91      | 0.82           | 0.21      |
| 17体・転写100%      | 1.72        | 1.27       | 1.15        | 0.87        | 0.94        | 0.94      | <b>0.87</b>    | 0.23      |
| 公開素材・転写0%       | 1.08        | 0.70       | 0.56        | 0.44        | 0.42        | 0.41      | <b>0.38</b>    | 0.15      |
| 公開素材・転写100%     | 0.96        | 0.65       | 0.49        | 0.54        | 0.42        | 0.42      | <b>0.39</b>    | 0.16      |

## 5.5 公開素材での評価

手法の設計と定数の選択に使っていない公開素材の17枚（5.1節）で評価した結果を、表3と表4の最後の2行に示す。境界項の倍率、グラフカットの $\lambda$ 、手がかりの重みを含め、すべての定数は他のデータセットで定めた値のままである。場面数は転写0%と100%のそれぞれで102、試行数は204である。

楕円をそのまま使うとIoUは転写0%で0.600、100%で0.596になる。適応（色）では0.912と0.874になり、事前分布のみでは0.825と0.775である。正解スコアを用いると、上限は0.978と0.962になる。手がかりの構成のうち、IoUが最大なのは、転写0%で適応（色とパッチ）(0.917)、100%で適応（色とパッチ）(0.874)である。固定重みと最良の構成のIoUの差は、転写0%で0.006、100%で0.001にとどまる。

構成どうしの差は、場面を単位とする信頼区間とあわせて次のとおりである。固定重みのテクスチャのみに対するIoUの差は、転写0%で+0.032 [+0.020, +0.044]、100%で+0.012 [+0.002, +0.022]である。固定重みの色とパッチのみに対する差は、転写0%で+0.025 [+0.009, +0.041]、100%で+0.077 [+0.057, +0.095]である。適応（色）の固定重みに対する差は、転写0%で+0.001 [-0.005, +0.007]、100%で+0.001 [-0.002, +0.004]で、適応（色とパッチ）では+0.006 [-0.001, +0.013]と+0.001 [-0.003, +0.005]である。

固定重みは、テクスチャのみと色とパッチのみのどちらに対しても、転写0%と100%の両方で、信頼区間が0を含まない差で上回る。

画素単位のグラフカットと木の切断（適応（色）のスコア）の差は、IoUで転写0%が+0.007 [+0.003, +0.012]、100%が+0.011 [+0.003, +0.018]、MAEで転写0%が-0.052 [-0.064, -0.040]、100%が-0.071 [-0.087, -0.056]である。差はグラフカットから木の切断を引いた値で、転写0%と100%のどちらでも、グラフカットのIoUが高く、MAEが小さい。

**表5** 公開素材での前景のIoUとMAE（精緻化後。各列の最良を太字で示す。正解スコアは上限として別に示し、太字の対象に含めない）。

| 構成        | 転写0%         |             | 転写100%       |             |
|-----------|--------------|-------------|--------------|-------------|
|           | IoU          | MAE         | IoU          | MAE         |
| 楕円をそのまま使用 | 0.600        | 1.08        | 0.596        | 0.96        |
| 事前分布のみ    | 0.825        | 0.70        | 0.775        | 0.65        |
| テクスチャのみ   | 0.879        | 0.56        | 0.861        | 0.49        |
| 色とパッチのみ   | 0.885        | 0.44        | 0.796        | 0.54        |
| 固定重み      | 0.911        | 0.42        | 0.873        | 0.42        |
| 適応（色）     | 0.912        | 0.41        | 0.874        | 0.42        |
| 適応（色とパッチ） | <b>0.917</b> | <b>0.38</b> | <b>0.874</b> | <b>0.39</b> |
| 正解スコア     | 0.978        | 0.15        | 0.962        | 0.16        |

## 5.6 色転写の量と手がかり

色転写の量を変えたときの、手がかりの性質と結果を図7に示す。星雲の背景に対する立ち絵1の5条件である。

楕円の内側と外側の分布のバタチャリヤ係数は、色で、転写0%の0.02から100%の0.59に増え、分布が重なっていく。テクスチャの係数は0.13から0.21でほとんど変わらず、パッチの係数は0.00から0.07と小さいままである。楕円内の前景と背景を分けるAUC（正解から求めた値）は、色が0.99から0.74に、パッチが0.99から0.93に下がるのに対し、テクスチャは0.98から0.97でほぼ一定である。色の係数の増加は、色のAUCの低下と対応する。式(16)で定めた色の重みは、0.37から0.00に下がる。

前景のIoUは、色とパッチのみでは0.960から0.815に下がるのに対し、テクスチャのみは0.965から0.961への低下にとどまる。固定重みのテクスチャのみに対するIoUの差は、転写0%、25%、50%、75%、100%の順に0.000、0.000、+0.003、-0.001、+0.001で、いずれも絶対値が0.003以下で、同程度である。適応(色)の固定重みに対する差は、転写100%で-0.003、テクスチャのみに対する差は-0.002である。

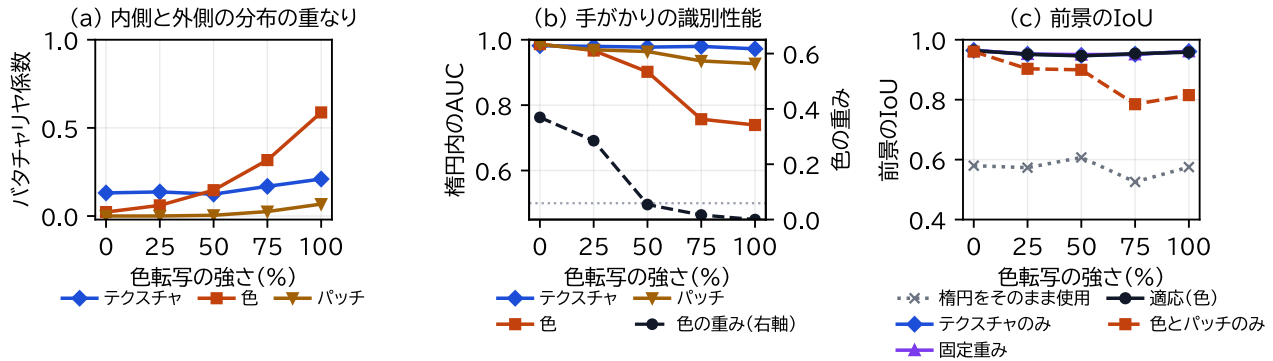


図7 色転写の量に対する、(a)内側と外側の分布のバタチャリヤ係数、(b)手がかりの識別性能（楕円内のAUC）と色の重み、(c)前景のIoU（立ち絵1、星雲、各8試行）。

## 5.7 信頼度の推定

全試行について、バタチャリヤ係数と楕円内のAUCの関係を図8に示す。係数が小さいほどAUCが高い関係が、テクスチャと色で見られる。立ち絵1、立ち絵2、17体の全584試行の順位相関（Spearman）は、テクスチャで-0.73、色で-0.69である。データセットごとの順位相関は、テクスチャでは立ち絵1で-0.74、立ち絵2で-0.46、17体の転写0%で-0.51、17体の転写100%で-0.62、色では立ち絵1で-0.61、立ち絵2で-0.53、17体の転写0%で-0.53、17体の転写100%で-0.41である。一方、パッチの係数は81%の試行で0.1未満に集まり、順位相関は-0.22と小さい。パッチの係数は識別性能の差を区別できないため、重みの調整には使わない。

データセットごとの楕円内の平均AUCは、17体の転写100%でテクスチャが0.78、色が0.67、パッチが0.82であり、色の識別性能が低くても、パッチはテクスチャを上回る。立ち絵1では、テクスチャが0.91、色が0.83、パッチが0.95で、パッチが最も高い。

係数は、手がかりの種類の間では直接比べにくい。色とテクスチャの係数の大小と、AUCの大小が食い違う試行は、584試行の25%を占める。星雲の背景で転写100%の条件では、色の係数が0.59、色のAUCが0.74であるのに対し、テクスチャの係数は0.21、AUCは0.97である。係数は、楕円の内側と外側という、境界から離れた画素の標本どうしの分布の離れを測る。境界付近の背景の分布は、外側の標本とは異なることがある。標本を取る領域と判定する領域とで特徴量の分布が異なるこの状況は、共変量シフト(covariate shift)と呼ばれる[43]。その大きさは、手がかりの種類ごとに異なると考えられる。このため、色の重みは、基準とするテクスチャの係数との比で定める（式(16)）。

楕円の内側と外側を、前景と背景の誤りを含むラベルとみなすと、この関係の理論的な背景は次のとおりである。クラスごとに一定の割合でラベルが入れ替わる場合、誤ったラベルに対するAUCは、正しいラベルに対するAUCの一次関数になる[29]。したがって、同じ標本を使う二つの手がかりの識別性能の順序は、誤ったラベルのもとでも保たれる。実際には、入れ替わりが画素の位置や外観に依存するため、この

仮定は成り立たない。図8の順位相関の絶対値も、テクスチャで0.73、色で0.69であり、単調な関係を表す1には届かない。

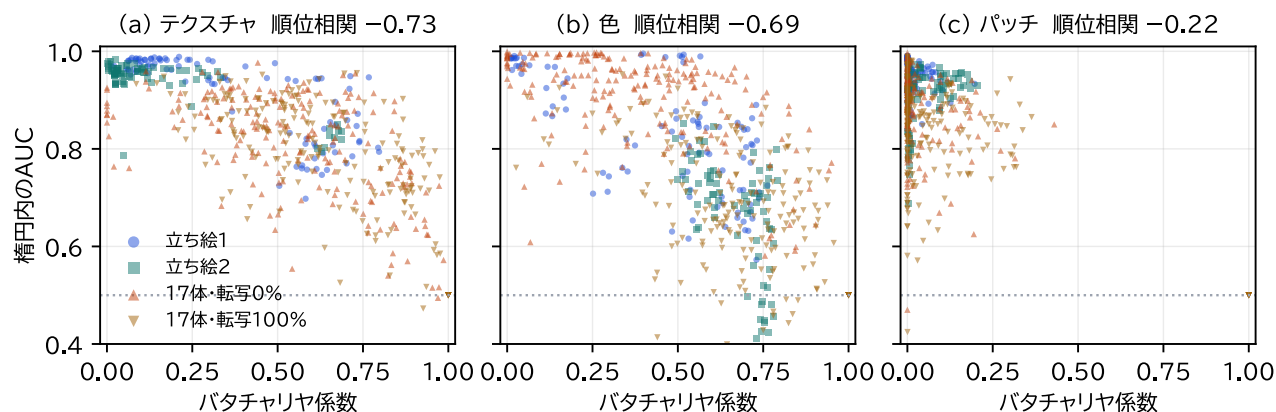


図8 バタチャリヤ係数と楕円内のAUC (1点が1試行、立ち絵1、立ち絵2、17体の全試行)。係数は正解を使わずに求め、AUCは正解から求めた。(c)のパッチの係数は0付近に集まる。

## 5.8 境界項の形と強さ

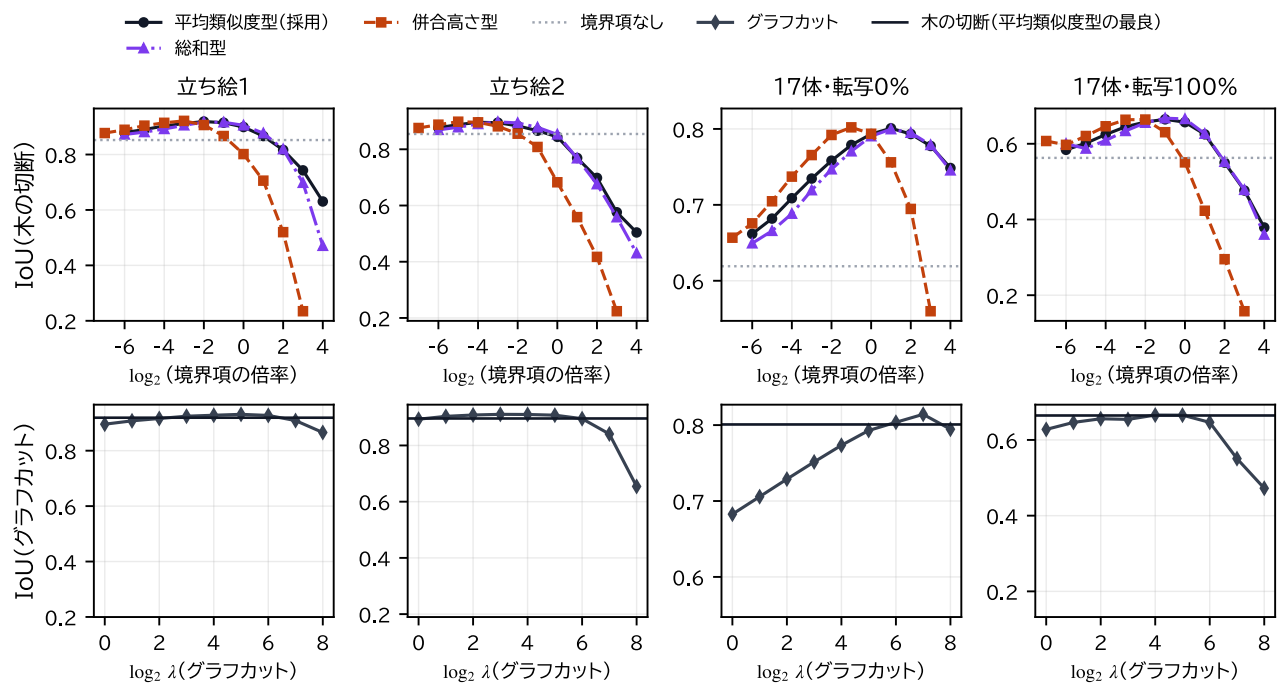
境界項の三つの形(4.5節)を、同じスコア(適応(色))のもとで比べる(表6、図9上段)。倍率は形ごとに、開発用の画像の平均IoUが最大になる値を選んだ。選んだ倍率での開発用の画像の平均IoUは、平均類似度型が0.786、総和型が0.787、併合高さ型が0.787で、差は0.002以内である。境界項の強さを調整すれば、三つの形の精度は同程度になる。

差は倍率への感度に現れる。選んだ倍率を4倍にすると、IoUは併合高さ型で0.072、総和型で0.065、平均類似度型で0.022下がり、併合高さ型の低下が最も大きい。1/4倍にすると、総和型で0.014、平均類似度型で0.022、併合高さ型で0.022下がり、形による差は0.009以内である。併合高さ型は、境界辺のうち最も類似度が高い一本の類似度を使う。この値は境界辺の類似度の平均以上なので、同じ倍率でもペナルティが強くなり、最適な倍率が小さい(併合高さ型が0.25、平均類似度型が0.5、総和型が1)。正解スコアを用いたときのIoUは、総和型が0.915、平均類似度型が0.925、併合高さ型が0.910であり、平均類似度型が正解の形を最もよく保つ。

最適な倍率はデータセットによって異なる。平均類似度型では、立ち絵1が0.25、立ち絵2が0.0625、17体の転写0%が2、転写100%が0.5で最大になり、最大で32倍の差がある。倍率を0.5の一つに決めると、各データセットの最良に対するIoUの低下は、立ち絵1で0.004、立ち絵2で0.030、17体の転写0%で0.022、転写100%で0.000である。境界項をなくすと、IoUは0.852、0.854、0.619、0.563に下がり、いずれのデータセットでも境界項は有効である。

**表6** 境界項の形ごとの切断のIoU（精緻化前の二値マスク）。倍率は開発用の画像（立ち絵1と17体）の平均を最大にするものを形ごとに選んだ。右の三列は、開発用の画像の平均について、選んだ倍率の1/4倍と4倍のときの変化、および正解スコアを用いたときのIoUである。

| 境界項の形      | 選んだ倍率 | 立ち絵1  | 立ち絵2  | 17体・転写0% | 17体・転写100% | 開発用の画像の平均 | 倍率を1/4倍 | 倍率を4倍  | 正解スコア |
|------------|-------|-------|-------|----------|------------|-----------|---------|--------|-------|
| 平均類似度型（採用） | 0.5   | 0.915 | 0.866 | 0.779    | 0.665      | 0.786     | -0.022  | -0.022 | 0.925 |
| 総和型        | 1     | 0.905 | 0.853 | 0.790    | 0.666      | 0.787     | -0.014  | -0.065 | 0.915 |
| 併合高さ型      | 0.25  | 0.906 | 0.855 | 0.792    | 0.664      | 0.787     | -0.022  | -0.072 | 0.910 |



**図9** 境界項の強さに対するIoU（精緻化前の二値マスク、適応（色）のスコア）。上段は木の切断の三つの形、下段はグラフカットで、点線は境界項なしのIoU、水平線は平均類似度型の最良のIoUである。

## 5.9 グラフカットとの比較

同じスコア  $u$ 、同じ辺の類似度  $\omega_e$ 、同じ精緻化のもとで、画素単位のグラフカットと比べる。グラフカットは、前景の画素の集合を  $F$  として、 $-\sum_{p \in F} u(p) + \lambda \sum_{(p,q): p \in F, q \notin F} \omega_{pq}$  を、画素の4近傍のグラフ上で最小カットにより最小にする[3]。 $\lambda$  は、開発用の画像の平均IoUが最大になる32とした（図9下段）。

表7に結果を示す。グラフカットのIoUは、立ち絵2、公開素材の転写0%、公開素材の転写100%で、信頼区間が0を含まない差で木の切断より高い（差は順に+0.024、+0.007、+0.011）。木の切断のほうが高いのは、17体の転写100%である（差は-0.014）。立ち絵1、17体の転写0%では、差の信頼区間が0を含む。MAEの差は、六つのデータセットの順に-0.078、-0.122、-0.059、-0.034、-0.052、-0.071で、いずれもグラフカットのほうが小さく、信頼区間は0を含まない。木の切断が選べるのは、単連結の併合で作った領域だけであり、輪郭のコントラストが低い場所で背景の領域と細くつながっていると、その部分だけを分けられないことがある。色転写100%のデータセットどうしても、17体では木の切断が、公開素材ではグラ

フカットが高く、この向きは画像のデータセットに依存する。グラフカットの $\lambda$ も、データセットごとの最良の値が8から128まで変わる。

最適化1回の計算時間を、同じスコアのもとで測った。幅256の画像 ( $N$  は3.3万~3.7万) の5場面について、計40個のスコアを解いた。木の動的計画法 (C#, 単一スレッド、スコアの合計から切断の復元まで) は1回あたり中央値1.10 ミリ秒、同じ画素グラフに対するDinic法による最小カット (C#) は47.8 ミリ秒で、約44倍である。どちらも、木とグラフの構築と、1回目の実行を除いた、2回目の実行の時間を問題数で割った値である。測定は13th Gen Intel(R) Core(TM) i7-1360P (12コア)、.NET 10で行った。他の処理の影響で、測定のたびに値は変わる。Dinic法は単純な実装であり、画像の格子グラフでは、BoykovとKolmogorovのアルゴリズム[44]のほうが速い。同じ問題を、KolmogorovによるC++の実装を呼び出すPyMaxflow 1.3.2で解き、maxflowの呼び出しだけの時間を測ると、1回あたり中央値1.55 ミリ秒で、木の動的計画法の約1.4倍である。この値にはグラフの構築を含めず、実装の言語もC#ではなくC++であるため、木の動的計画法との比は目安にとどまる。木の構築と境界項の計算は画像ごとに1回で済み、楕円を動かす操作では、スコアの再計算のあとに動的計画法だけを繰り返す。

**表7** 木の切断とグラフカットの比較 (適応 (色) のスコア、精緻化後)。差はグラフカットから木の切断を引いた値で、場面を単位とするブートストラップ法で求めた。

| データセット            | 木の切断<br>IoU | グラフカット<br>IoU | IoUの差<br>[95%信頼区間]         | 木の切断<br>MAE | グラフカット<br>MAE | MAEの差<br>[95%信頼区間]      |
|-------------------|-------------|---------------|----------------------------|-------------|---------------|-------------------------|
| 立ち絵1・全12条件        | 0.926       | 0.934         | +0.007<br>[0.000, +0.015]  | 0.43        | 0.35          | -0.08<br>[-0.13, -0.03] |
| 立ち絵2・8背景・<br>100% | 0.888       | 0.912         | +0.024<br>[+0.008, +0.046] | 0.75        | 0.63          | -0.12<br>[-0.17, -0.07] |
| 17体・転写0%          | 0.789       | 0.794         | +0.005<br>[-0.001, +0.012] | 0.91        | 0.85          | -0.06<br>[-0.08, -0.04] |
| 17体・転写100%        | 0.681       | 0.667         | -0.014<br>[-0.026, -0.003] | 0.94        | 0.90          | -0.03<br>[-0.06, -0.01] |
| 公開素材・転写0%         | 0.912       | 0.919         | +0.007<br>[+0.003, +0.012] | 0.41        | 0.36          | -0.05<br>[-0.06, -0.04] |
| 公開素材・転写100%       | 0.874       | 0.885         | +0.011<br>[+0.003, +0.018] | 0.42        | 0.35          | -0.07<br>[-0.09, -0.06] |

## 5.10 解像度の影響

テクスチャの手がかりを、元の解像度の画像と、縮小した画像から計算したときの比較を表8に示す。入力を1/2 (幅640) に縮小すると、テクスチャのみのIoUは、立ち絵1で0.031、立ち絵2で0.051、17体の転写0%と100%で0.022と0.026下がる。1/4 (幅320) では、それぞれ0.077、0.130、0.058、0.064下がる。色とパッチを併用した場合も、1/2への縮小で0.022、0.040、0.018、0.023下がる。平坦な領域は縮小の平均で失われるため、テクスチャの手がかりは元の解像度の画素から計算する。

縮小画像の幅  $L$  (4.1節) を変えた結果を表9に示す。 $L = 128$  から256に上げると、テクスチャのみのIoUは、立ち絵1で+0.021、17体の転写0%と100%で+0.016と+0.025変わり、立ち絵2では+0.010変わる。256から384への変化は、立ち絵1で-0.004、立ち絵2で0.000、17体で+0.004と-0.001である。変化の絶対値は0.005以下で、幅を256より大きくしても、IoUはほとんど変わらない。

表8 テクスチャの手がかりを計算する入力画像の解像度と、切断のIoU（精緻化前の二値マスク）。「併用」は、テクスチャの重みを1.0、色とパッチの重みを適応（色）の値にした構成である。

| テクスチャの入力     | テクスチャのみ |       |          |            | 併用    |       |          |            |
|--------------|---------|-------|----------|------------|-------|-------|----------|------------|
|              | 立ち絵1    | 立ち絵2  | 17体・転写0% | 17体・転写100% | 立ち絵1  | 立ち絵2  | 17体・転写0% | 17体・転写100% |
| 元の解像度（幅1280） | 0.900   | 0.859 | 0.737    | 0.634      | 0.915 | 0.866 | 0.779    | 0.665      |
| 1/2に縮小（幅640） | 0.870   | 0.807 | 0.716    | 0.609      | 0.893 | 0.826 | 0.761    | 0.642      |
| 1/4に縮小（幅320） | 0.823   | 0.728 | 0.680    | 0.570      | 0.853 | 0.766 | 0.737    | 0.616      |

表9 縮小画像の幅  $L$  と、前景のIoU / MAE（テクスチャのみ、精緻化後）。17体のデータセットは3場面ごとに1場面で評価した。

| 縮小画像の幅 $L$ | 立ち絵1         | 立ち絵2         | 17体・転写0%     | 17体・転写100%   |
|------------|--------------|--------------|--------------|--------------|
| 128        | 0.892 / 0.46 | 0.872 / 0.68 | 0.725 / 1.30 | 0.628 / 1.32 |
| 256        | 0.913 / 0.49 | 0.882 / 0.81 | 0.741 / 1.29 | 0.653 / 1.28 |
| 384        | 0.909 / 0.55 | 0.882 / 0.87 | 0.745 / 1.26 | 0.651 / 1.27 |

### 5.11 事前分布の内側の強さ

事前分布の内側の縮小率  $\gamma$ （4.4節）を変えた結果を表10に示す。立ち絵1では  $\gamma = 0.5$ 、立ち絵2では  $\gamma = 0.7$  で最大になるのに対し、17体では、転写0%が  $\gamma = 0$ 、転写100%が  $\gamma = 0.15$  で最大になり、内側の正のスコアを弱めるほどIoUが高い。 $\gamma = 0.35$  と最良の差は、立ち絵1で0.000、立ち絵2で0.005、17体の転写0%で0.030、転写100%で0.008である。誤差モデルMで楕円の内側を前景が占める割合は、立ち絵では約0.61、17体では約0.42であり（3.3節）、内側の割合が小さいデータセットほど、内側の事前分布を弱めたときのIoUが高い。 $\gamma$  もまた、データセットにより最適値が異なる定数である。

表10 事前分布の内側の縮小率  $\gamma$  を変えたときの切断のIoU（精緻化前の二値マスク、適応（色）のスコア）。各列の最大を太字で示す。

| 内側の縮小率 $\gamma$ | 立ち絵1         | 立ち絵2         | 17体・転写0%     | 17体・転写100%   | 開発用の画像の平均    |
|-----------------|--------------|--------------|--------------|--------------|--------------|
| 1               | 0.901        | 0.868        | 0.712        | 0.613        | 0.742        |
| 0.7             | 0.913        | <b>0.871</b> | 0.739        | 0.636        | 0.763        |
| 0.5             | <b>0.915</b> | 0.869        | 0.764        | 0.651        | 0.777        |
| 0.35            | 0.915        | 0.866        | 0.779        | 0.665        | 0.786        |
| 0.25            | 0.908        | 0.862        | 0.789        | 0.671        | 0.789        |
| 0.15            | 0.906        | 0.858        | 0.797        | <b>0.672</b> | <b>0.792</b> |
| 0               | 0.898        | 0.855        | <b>0.808</b> | 0.671        | 0.792        |

## 5.12 精緻化

精緻化の各段を表11に示す。二値マスクを双一次に拡大する場合に対し、ジョイントバイラテラルアップサンプリングによるMAEの変化は、立ち絵1で-0.03、立ち絵2で-0.01、17体の転写0%と100%で-0.01と0.00である。全画素に色の補正をかけると、MAEは立ち絵1で0.89、17体の転写0%で1.49と、補正をかけない場合の0.44と0.95より大きくなる。補正の対象を境界付近の帯に限ると、MAEは立ち絵1で0.43、17体の転写0%で0.94と、補正をかけない場合と同程度になる。ガイドッドフィルタを加えたときのIoUの変化は、立ち絵1で+0.008、立ち絵2で+0.013、17体の転写0%と100%で+0.004と+0.010である。

**表11** 精緻化の各段の効果（前景のIoU / MAE）。切断は適応（色）のスコアを使う木の切断で、立ち絵1、立ち絵2、17体の全試行から求めた。

| 精緻化                    | 立ち絵1         | 立ち絵2         | 17体・転写0%     | 17体・転写100%   |
|------------------------|--------------|--------------|--------------|--------------|
| 二値マスクの双一次拡大            | 0.912 / 0.46 | 0.867 / 0.78 | 0.774 / 0.96 | 0.664 / 0.97 |
| ジョイントバイラテラルアップサンプリング   | 0.917 / 0.44 | 0.872 / 0.77 | 0.780 / 0.95 | 0.669 / 0.98 |
| +色の補正（全画素）             | 0.910 / 0.89 | 0.875 / 1.24 | 0.785 / 1.49 | 0.669 / 1.33 |
| +色の補正（帯の中だけ）           | 0.919 / 0.43 | 0.875 / 0.75 | 0.784 / 0.94 | 0.671 / 0.96 |
| +ガイドッドフィルタ（補正なし）       | 0.925 / 0.44 | 0.888 / 0.77 | 0.786 / 0.92 | 0.680 / 0.95 |
| +色の補正（帯）+ガイドッドフィルタ（本稿） | 0.926 / 0.43 | 0.888 / 0.75 | 0.789 / 0.91 | 0.681 / 0.94 |

## 5.13 楕円の誤差の大きさ

楕円の誤差モデルを変えた結果を表12に示す（3.3節）。立ち絵1、立ち絵2、17体の四つのデータセットと四つの誤差モデルの16通りのうち15通りで、テクスチャのみ、事前分布のみ、楕円をそのまま使う場合の順にIoUが高い。例外は、立ち絵2の誤差モデルS（テクスチャのみ0.893、事前分布のみ0.675、楕円をそのまま使う場合0.676）である。テクスチャの手がかりによる事前分布のみに対する向上は、+0.051から+0.253の範囲に収まる。誤差が小さいモデルTから大きいモデルLに変えたときのテクスチャのみのIoUの変化は、立ち絵1で-0.051、立ち絵2で-0.054、17体の転写0%と100%で-0.120と-0.122である。

**表12** 楕円の誤差モデルを変えたときの前景のIoU / MAE（精緻化後）。誤差モデルMの値は主な結果と同じ試行である。

| データセット     | 誤差モデル | 楕円をそのまま      | 事前分布のみ       | テクスチャのみ      | 正解スコア        |
|------------|-------|--------------|--------------|--------------|--------------|
| 立ち絵1       | M     | 0.575 / 1.33 | 0.760 / 1.06 | 0.913 / 0.49 | 0.966 / 0.24 |
|            | S     | 0.625 / 1.21 | 0.688 / 1.22 | 0.894 / 0.50 | 0.966 / 0.24 |
|            | T     | 0.688 / 0.96 | 0.775 / 0.98 | 0.927 / 0.42 | 0.966 / 0.24 |
|            | L     | 0.499 / 1.73 | 0.672 / 1.34 | 0.875 / 0.70 | 0.966 / 0.24 |
| 立ち絵2       | M     | 0.588 / 1.67 | 0.689 / 1.36 | 0.882 / 0.81 | 0.944 / 0.30 |
|            | S     | 0.676 / 1.20 | 0.675 / 1.22 | 0.893 / 0.63 | 0.944 / 0.30 |
|            | T     | 0.746 / 0.89 | 0.774 / 0.90 | 0.905 / 0.54 | 0.944 / 0.30 |
|            | L     | 0.499 / 2.26 | 0.597 / 1.81 | 0.851 / 1.04 | 0.944 / 0.30 |
| 17体・転写0%   | M     | 0.406 / 1.87 | 0.692 / 1.28 | 0.750 / 1.14 | 0.961 / 0.21 |
|            | S     | 0.486 / 1.55 | 0.706 / 1.19 | 0.787 / 0.90 | 0.961 / 0.21 |
|            | T     | 0.516 / 1.30 | 0.759 / 0.93 | 0.811 / 0.79 | 0.961 / 0.21 |
|            | L     | 0.343 / 2.44 | 0.620 / 1.73 | 0.690 / 1.58 | 0.961 / 0.21 |
| 17体・転写100% | M     | 0.404 / 1.72 | 0.565 / 1.27 | 0.651 / 1.15 | 0.870 / 0.23 |
|            | S     | 0.482 / 1.28 | 0.579 / 1.05 | 0.676 / 0.89 | 0.870 / 0.23 |
|            | T     | 0.512 / 1.14 | 0.624 / 0.90 | 0.705 / 0.81 | 0.870 / 0.23 |
|            | L     | 0.341 / 2.24 | 0.497 / 1.69 | 0.583 / 1.58 | 0.870 / 0.23 |

#### 5.14 定性的な結果

図10に、条件ごとにIoUが中央値に近い試行を一つずつ示す（立ち絵1は星雲とそれ以外の背景から一つずつ、公開素材は転写0%と100%の試行と、転写100%でIoUが下位1/8にあたる試行）。白は正解と一致した前景、赤は背景を前景とした誤り、青は前景を背景とした誤りである。楕円をそのまま使うと、楕円に収まらない髪や衣服の端が青の誤りに、楕円の内側の背景が赤の誤りになる。事前分布のみでは、デンドログラムの境界に沿って形が整うが、前景の一部が背景になり、背景の一部が前景に残る。立ち絵1の二つの例と立ち絵2の例では、テクスチャのみ、固定重み、適応（色）で輪郭の多くが正解と一致し、誤りは耳や髪の手先、衣服の内部の小さな領域に残る。立ち絵1の夜景の例では、テクスチャのみが足元の背景の領域を前景に含める。公開素材の転写0%の例でも輪郭の多くが一致し、誤りは髪の手先に残る。公開素材の転写100%の二つの例では、背景の平坦な領域（壁の帯）が前景に入る構成が多く、平坦さを使うテクスチャのみで特に目立つ。グラフカットは、立ち絵1の星雲の例、立ち絵2の例、公開素材の転写0%の例、公開素材の転写100%の例でIoUが最大であるが、立ち絵1の夜景の例では固定重みの0.93より低く、公開素材の下位の例では固定重みの0.86より低い。

図11に、同じ試行のうち二つの背景のぼかし描画と、正解の $\alpha$ による正解描画との差を示す。立ち絵1の例では、楕円をそのまま使う、あるいは事前分布のみでは、前景の縁が背景と一緒にぼける、あるいは背景の色が前景に残る誤りが縁に沿って目立ち、テクスチャのみと適応（色）では、差は縁の細い範囲にと

どまる。公開素材の転写0%の例では、楕円をそのまま使うと、楕円の内側の背景のテクスチャがぼけずに残る。テクスチャのみと適応（色）では、白い髪の毛の先が背景として扱われて、差が髪に集まる。MAEが最も小さいのは、立ち絵1の例で適応（色）の0.17、公開素材の例で事前分布のみの0.31である。

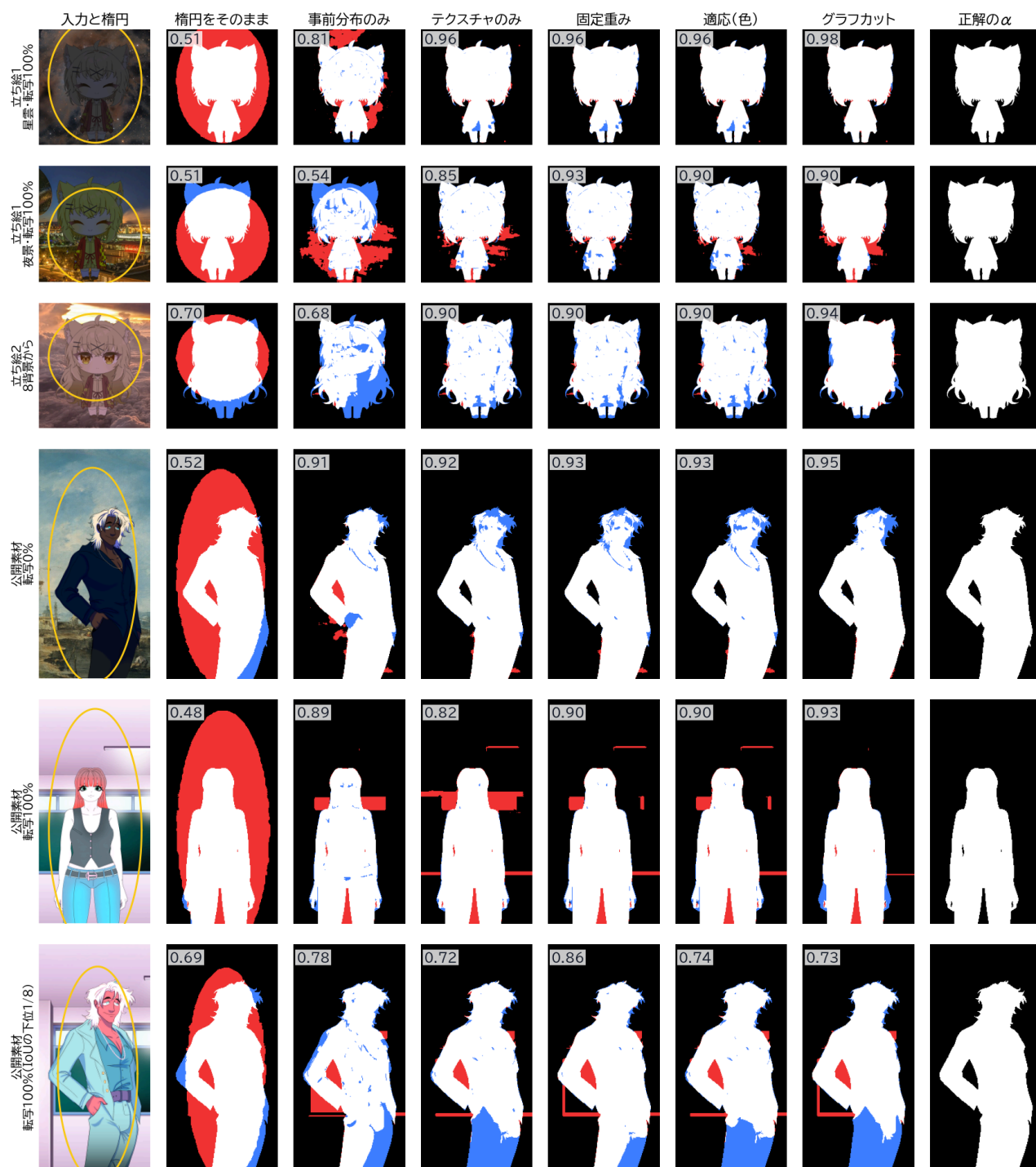


図10 条件ごとのIoUが中央値に近い試行の結果。左から、入力と指定した楕円、楕円をそのまま使った結果、事前分布のみ、テクスチャのみ、固定重み、適応（色）、グラフカット、正解の $\alpha$ を並べた。白は正解と一致した前景、赤は背景を前景とした誤り、青は前景を背景とした誤りで、左上の数値はIoUである。最後の行は、公開素材の転写100%でIoUが下位1/8にあたる試行である。

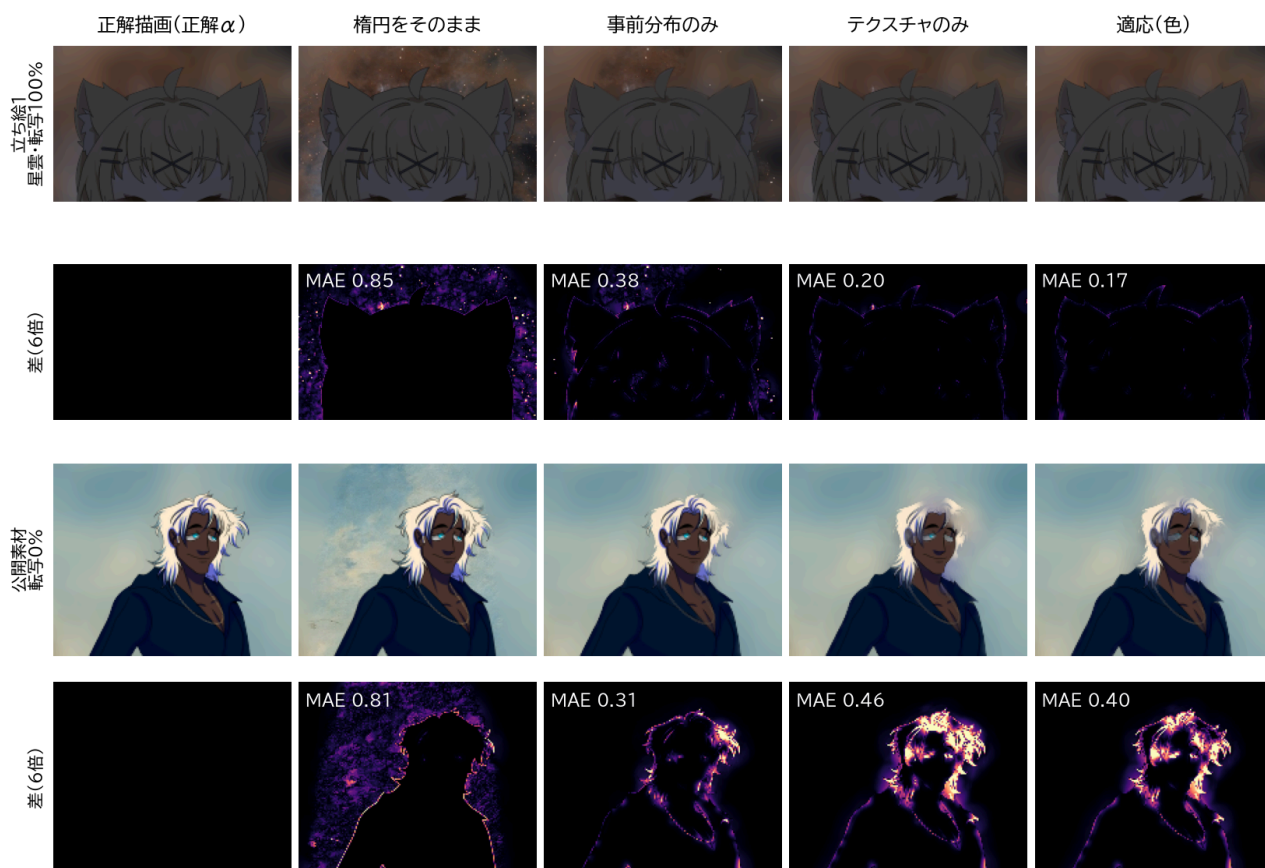


図 11 背景のぼかし描画（上段）と、正解描画との差の絶対値を6倍にした画像（下段）。正解描画は正解の $\alpha$ による描画である。

## 6 考察

### 6.1 結果の読み方

手がかりの有効性はデータセットによって異なる。色とパッチの手がかりは、楕円の内側と外側の標本の分布の差を使う。この差が境界付近の画素にも当てはまるのは、前景と背景の絵柄が異なり、背景の標本が境界付近の背景と同じ分布に従うと見なせる場合である。立ち絵1は、輪郭線と一様な塗りからなる前景と、写真や絵画の背景の組である。色転写は前景の色の平均と分散を背景に合わせるため、この絵柄の違いのうち色の分布の部分弱める。一方、一様な塗りは変換後も一様なので、テクスチャの手がかりは保たれる。実際、立ち絵1の楕円内の平均AUCは、テクスチャが0.91、色が0.83、パッチが0.95である。色転写の強さを0%から100%に変えると、色のAUCは0.99から0.74に下がるが、テクスチャのAUCは0.98から0.97にとどまる。17体のデータセットでは、前景の絵柄が多様で陰影や細かい模様を持つものが多く、テクスチャの平均AUCは転写0%で0.81、100%で0.78にとどまる。これに対し、色は転写0%で0.91、パッチは100%でも0.82と、テクスチャより高い。この違いが、テクスチャのみと色とパッチのみの優劣がデータセットで変わる主な理由と考えられる。

固定重みはデータセットを問わず安定している。六つのデータセットのすべてで、固定重みと最良の構成のIoUの差は0.012以内である。これに対し、テクスチャのみでは最大0.049、色とパッチのみでは最大

0.102の差が生じる。画像のデータセットが事前にわからない場合は、手がかりを併用する固定重みが妥当な選択である。

信頼度に基づく重みの効果は小さい。色の重みを調整する適応（色）の効果は、固定重みに対して最大でも0.004で、符号はデータセットにより異なる。パッチの重みも調整すると、17体の転写0%、17体の転写100%では信頼区間が0を含まない向上を示し、立ち絵1、立ち絵2、公開素材の転写0%、公開素材の転写100%では差の信頼区間が0を含む。効果が小さい理由は、次のように考えられる。係数は境界から離れた標本の分布の離れであり、境界付近の画素の識別性能とは一致しない。また、色の重みを調整できる範囲は0から0.5と狭く、調整後もテクスチャとパッチの手がかりが残る。パッチの重みについては、係数が識別性能との対応が弱く、調整の根拠として使いにくい。係数と手がかりの識別性能の順位相関の絶対値は、テクスチャで0.73、色で0.69である。この関係を境界付近の画素の識別性能と対応づける方法は、本稿では示していない。

木の切断とグラフカットは、精度と最適化の速さで異なる。グラフカットから木の切断を引いたIoUの差は、立ち絵1で+0.007、立ち絵2で+0.024、17体の転写0%で+0.005、17体の転写100%で-0.014、公開素材の転写0%で+0.007、公開素材の転写100%で+0.011である。信頼区間が0を含まず、グラフカットが高いデータセットは立ち絵2、公開素材の転写0%、公開素材の転写100%、木の切断が高いデータセットは17体の転写100%である。ぼかし描画のMAEは、六つのデータセットのすべてでグラフカットのほうが小さい。最適化1回の計算時間は、木の動的計画法がDinic法による最小カットの約44分の1、BoykovとKolmogorovのアルゴリズムの約1.4分の1である（5.9節）。木の切断は、画像のデンドログラムに含まれる領域しか表現できず、細い連結で背景とつながる細部を分けられないことがある。この制約が、グラフカットが高くなるデータセットの一因と考えられる。17体の転写100%で木の切断が高かった理由は、本稿の実験からは特定できない。楕円を動かすたびに切断を更新する対話的な処理では、最適化の時間が短い木の切断が適すると考えられる。輪郭が明瞭な画像で最終的な描画の精度を優先する場合は、グラフカットが適すると考えられる。

## 6.2 妥当性への脅威

合成による評価。評価に使った画像は、立ち絵と背景を合成したものである。前景と背景は別の出自を持つため、実際の写真に比べて絵柄の違いが大きく、外観の手がかりに有利に働くと考えられる。写真の人物や、前景と背景が連続する画像での結果は本稿では示していない。また、元の解像度の平坦さの手がかりは、一様な塗りの前景を前提にしており、撮像のノイズを持つ前景に対して働くかどうかは確かめていない。

評価指標。IoUは、0.5で二値化した前景の形の一致を測る。MAEは、ぼかし描画への視覚的な影響を測る。どちらも、髪や半透明の部分のアルファの中間値の精度を直接には測らない。マッピングの評価で使われるSADなどの誤差尺度[24]は報告していない。

信頼区間の単位。信頼区間は、場面を単位とする復元抽出で求めた。17体と公開素材では、同じ立ち絵から6場面を作っており、それらは互いに似ているため、区間が狭くなっている可能性がある。立ち絵1と立ち絵2はキャラクターが1体なので、立ち絵を単位にした区間は求められない。グラフカットと木の切断のIoUの差について、立ち絵を単位に再標本化すると、公開素材の転写100%で+0.011 [-0.001, +0.021] となり、区間は0を含む。

開発用のデータセットと定数の選択。立ち絵1と17体は、手法の設計と定数の選択に使った。選択した定数のうち、境界項の倍率とグラフカットの $\lambda$ は、最適値がデータセットにより大きく異なる。既定の事前分布の内側の縮小率も、最適値がデータセットにより異なる。設計に使っていない立ち絵2と公開素材でも、色とパッチのみが固定重みを下回り、固定重みと最良の構成の差が0.006以内であるという傾向は保たれた。立ち絵2は立ち絵1と同じキャラクターの髪型違いで、絵柄が共通するため、立ち絵1に対する独立性は低い。背景も8種にすぎない。公開素材の立ち絵17枚のうち12枚は同じ作者の7セットで、絵柄が近い。結果の一般性は限られる。17体のデータセットは、本稿の実験の前に行った探索でも使っており、その影響を完全には切り離せない。

楕円の誤差モデル。利用者の楕円の誤差は、正解の外接矩形から作った楕円に、倍率と中心のずれを一樣乱数で加える模擬である。実際の操作で生じる誤差の分布や、系統的な偏りとは異なる可能性がある。事前分布の内側の縮小率は、誤差モデルMでの楕円の内側の割合に依存する（5.11節）。誤差モデルを変えた結果は5.13節に示したが、実際の利用者の操作は評価していない。

描画の簡略化。ぼかし描画は、円板の点拡がり関数による均一なぼかしである。深度に応じた変化や、明るい光源の広がりとは扱っておらず、MAEはこの描画に対する値である。

外部の手法との比較。本稿の比較は、構成どうしと、同じスコアを使う画素単位のグラフカットに限られる。GrabCutなどの既存の手法とは比較していない。学習済みの分割モデル[34]とも比較しておらず、これらのモデルは、本稿の手法より高い精度を出す可能性がある。

計算時間の測定範囲。計算時間は、動的計画法と最小カットの1回あたりの時間だけを測った。楕円を動かすたびに必要の手がかりの再計算は測っておらず、特にパッチの最近傍探索は計算量が他の段より大きい（4.9節）。最適化の速さの差が、手がかりの再計算を含む更新全体の時間にどれだけ影響するかは、本稿では示していない。

### 6.3 今後の課題

三つの課題が残る。第一は、写真の人物を含む画像での評価である。第二は、手がかりの信頼度を、境界付近の画素の識別性能と対応づける方法である。そのためには、楕円の内側と外側の標本のうち、境界に近い画素だけを使う方法や、手がかり間の一致を使う方法が考えられる。第三は、木の切断とグラフカットの併用である。木の切断の結果を初期値として、境界付近の帯の中だけを画素単位で再判定すれば、計算時間を抑えながら、細部の精度を上げられる可能性がある。

## 7 おわりに

楕円で大まかに指定された前景を、単連結法のデンドログラムの上の最適な切断として推定する方法を述べ、手がかり、境界項、最適化のそれぞれの効果を、六つの画像データセットで評価した。

デンドログラムの上の切断により、楕円をそのまま使う場合のIoUを、0.40~0.60から、適応（色）の構成で0.68~0.93に上げた。元の解像度の平坦さを使うテクスチャの手がかりは、色転写で色とパッチの手がかりの識別性能が下がる合成画像で有効であり、立ち絵1の全12条件でIoUが0.913になった。一方、17体の転写0%では、色とパッチのみがテクスチャのみを上回る。両者の優劣はデータセットにより異なるが、併用する固定重みは、設計に使っていないデータセットを含むすべてのデータセットで、最良との差を0.012以内に抑えた。

楕円の内側と外側の分布のバタチャリヤ係数は、テクスチャと色の識別性能と対応した（順位相関 $-0.73$ 、 $-0.69$ ）。しかし、係数に基づく色の重みの調整が固定重みに与える効果は、IoUで最大 $0.004$ と小さかった。境界項の形は、強さを調整すれば精度が同程度になり、併合高さ型は倍率を大きくしたときの低下が最も大きかった。画素単位のグラフカットは、立ち絵と公開素材で木の切断より $0.007\sim 0.024$ 高いIoUを示し、木の切断は最適化の計算が速い。

本稿の評価は合成画像に限られ、学習済みモデルとの比較も行っていない。写真での評価、境界付近の画素に対応づけた信頼度、木の切断とグラフカットの併用が、今後の課題である。

## 謝辞

公開素材の評価には、次の素材を使用した。立ち絵はアルファチャンネル付きのPNGで、本稿では、背景との合成と色転写の加工を行った。背景は、いずれもCC0で公開されている画像を、16:9に切り出して拡大縮小して用いた。

### 立ち絵

- Breezy, “Visual Novel Sprites!”, CC0 1.0, <https://breezy-the-cat.itch.io/visual-novel-sprites>
- 4cher (Cyanea), “[FREE-TO-USE] Visual Novel Sprites”, Set 1, 2, 4, 5, 6, 7, CC BY 4.0, <https://iam-cyanea.itch.io/>（各セットのページはset1-vnspritesのように、末尾がset番号-vnspritesである）
- 4cher (Cyanea), “[FREE-TO-USE] Visual Novel Sprites”, Set 8, ライセンスの表示はなく、作者の説明文は商用利用を許しクレジットを不要としている, <https://iam-cyanea.itch.io/set8-vnsprites>

### 背景

- Wilfredor, “Quebec City night skylines (downtown 19).jpg”, CC0, [https://commons.wikimedia.org/wiki/File:Quebec\\_City\\_night\\_skylines\\_\(downtown\\_19\).jpg](https://commons.wikimedia.org/wiki/File:Quebec_City_night_skylines_(downtown_19).jpg)
- Anonymous0006, “Evening sunset skies.jpg”, CC0, [https://commons.wikimedia.org/wiki/File:Evening\\_sunset\\_skies.jpg](https://commons.wikimedia.org/wiki/File:Evening_sunset_skies.jpg)
- Ales Krivec aleskrivec, “Clouds mirrored in a mountain lake (Unsplash).jpg”, CC0, [https://commons.wikimedia.org/wiki/File:Clouds\\_mirrored\\_in\\_a\\_mountain\\_lake\\_\(Unsplash\).jpg](https://commons.wikimedia.org/wiki/File:Clouds_mirrored_in_a_mountain_lake_(Unsplash).jpg)
- A.-K. D., “Forest path along wall.jpg”, CC0, [https://commons.wikimedia.org/wiki/File:Forest\\_path\\_along\\_wall.jpg](https://commons.wikimedia.org/wiki/File:Forest_path_along_wall.jpg)
- Darko7411, “Iris Nebula.jpg”, CC0, [https://commons.wikimedia.org/wiki/File:Iris\\_Nebula.jpg](https://commons.wikimedia.org/wiki/File:Iris_Nebula.jpg)
- Mingiki, “Sunset rolling waves.jpg”, CC0, [https://commons.wikimedia.org/wiki/File:Sunset\\_rolling\\_waves.jpg](https://commons.wikimedia.org/wiki/File:Sunset_rolling_waves.jpg)
- Wilfredor, “Bike on a snowy street in Quebec City at night.jpg”, CC0, [https://commons.wikimedia.org/wiki/File:Bike\\_on\\_a\\_snowy\\_street\\_in\\_Quebec\\_City\\_at\\_night.jpg](https://commons.wikimedia.org/wiki/File:Bike_on_a_snowy_street_in_Quebec_City_at_night.jpg)
- Romain Guy from San Francisco, USA, “Mount Fuji at sunset, March 2025.jpg”, CC0, [https://commons.wikimedia.org/wiki/File:Mount\\_Fuji\\_at\\_sunset,\\_March\\_2025.jpg](https://commons.wikimedia.org/wiki/File:Mount_Fuji_at_sunset,_March_2025.jpg)
- Gaspard Dughet, “Imaginary Landscape MET DP225500.jpg”, CC0, [https://commons.wikimedia.org/wiki/File:Imaginary\\_Landscape\\_MET\\_DP225500.jpg](https://commons.wikimedia.org/wiki/File:Imaginary_Landscape_MET_DP225500.jpg)
- Francesco Guardi, “Fantastic Landscape MET DT9352.jpg”, CC0, [https://commons.wikimedia.org/wiki/File:Fantastic\\_Landscape\\_MET\\_DT9352.jpg](https://commons.wikimedia.org/wiki/File:Fantastic_Landscape_MET_DT9352.jpg)
- Théodore Rousseau, “A River Landscape MET DP274261.jpg”, CC0, [https://commons.wikimedia.org/wiki/File:A\\_River\\_Landscape\\_MET\\_DP274261.jpg](https://commons.wikimedia.org/wiki/File:A_River_Landscape_MET_DP274261.jpg)
- Robert Seldon Duncanson, “Landscape with Cows Watering in a Stream MET DT10837.jpg”, CC0, [https://commons.wikimedia.org/wiki/File:Landscape\\_with\\_Cows\\_Watering\\_in\\_a\\_Stream\\_MET\\_DT10837.jpg](https://commons.wikimedia.org/wiki/File:Landscape_with_Cows_Watering_in_a_Stream_MET_DT10837.jpg)
- Iletora, “School Hallway (morning)”, CC0, <https://iletora.itch.io/school-hallway>
- Iletora, “School Hallway (evening)”, CC0, <https://iletora.itch.io/school-hallway>

- Midnight68, “Classroom 002”, CC0, <https://opengameart.org/content/classroom-002>
- Midnight68, “Classroom 002”, CC0, <https://opengameart.org/content/classroom-002>

立ち絵1と立ち絵2は、筆者の同じキャラクターで、髪型だけが異なる。17体の立ち絵は第三者の著作物であり、数値の評価にだけ使用した。本稿には掲載していない。

本稿のコード、評価の結果、立ち絵1と立ち絵2の前景の色とアルファマップは、<https://github.com/routersys/EllipseForeground>で公開している。コードのライセンスはBSD 3-Clause Licenseである。立ち絵1と立ち絵2のデータは、定型のライセンスではなく独自の利用条件で提供し、AIの学習への利用、営利を目的とする利用、データだけの配布を禁止している。この条件は、本稿に載る立ち絵1と立ち絵2の図にも適用する。アルファマップの復元に使った書き出しは、背景を再配布できないため収録していない。

## 付録A 定数の一覧

本稿で用いた定数を表13に示す。「決め方」の列の「既定」は、本稿の実験の前に行った予備実験で定め、主な結果では変更していない値である。「選択」は、本稿の開発用の画像（立ち絵1と17体）の実験で、候補の中から選んだ値である。予備実験には、17体に近い構成の画像と人工的な合成画像を使っており、本稿の評価に使ったデータセットの一部を含む。立ち絵2と公開素材だけが、既定の値の決定にも、選択にも使っていないデータセットである。

表13 定数の一覧。

| 項目       | 定数              | 値                                                             | 決め方       |
|----------|-----------------|---------------------------------------------------------------|-----------|
| 縮小画像     | 幅 $L$           | 256                                                           | 既定（5.10節） |
|          | 範囲の正規化          | 2、98パーセントイル、下限0.04、上限1.5                                      | 既定        |
| 事前分布     | 境界の幅 $\kappa$   | 0.35                                                          | 既定        |
|          | 内側の縮小率 $\gamma$ | 0.35                                                          | 既定（5.11節） |
|          | 楕円の倍率 $a$       | 0.85、1、1.15                                                   | 既定        |
| 標本       | 中心側 $K$         | $\rho < 0.5$                                                  | 既定        |
|          | 外側 $B$          | $\rho > 1.3$                                                  | 既定        |
| 境界項      | 係数 $\lambda_0$  | 14/3                                                          | 既定        |
|          | 類似度の $\sigma$   | 0.05、0.1の平均                                                   | 既定        |
|          | 強さの倍率 $m$       | 0.5                                                           | 選択        |
|          | 解像度の補正          | $L / 128$                                                     | 既定        |
| 手がかり（共通） | 対数尤度比の打ち切り      | $\pm 2$ (ln 単位)、2で割る                                          | 既定        |
|          | スコアの範囲          | $[-1, 1]$                                                     | 既定        |
| 色        | ヒストグラム          | $8 \times 8 \times 8$ 、 $3 \times 3 \times 3$ で平滑化、 $10^{-5}$ | 既定        |
|          | 固定重みの値          | 0.25                                                          | 選択        |

| 項目        | 定数                       | 値                                                                   | 決め方 |
|-----------|--------------------------|---------------------------------------------------------------------|-----|
| パッチ       | 窓                        | 3×3 (27 次元)                                                         | 既定  |
|           | 標本数                      | 中心側3000 以下、外側は中心側の8 倍以下 (下限20)                                      | 既定  |
|           | 除外する範囲                   | $\ p - q\ _{\infty} \leq 2$                                         | 既定  |
|           | 係数 $\kappa_{\text{pat}}$ | 4                                                                   | 既定  |
|           | 固定重みの値                   | 0.3                                                                 | 選択  |
| テクスチャ     | 平坦の閾値                    | $m(p) \leq 2$                                                       | 既定  |
|           | 差の上限、平滑化                 | 32、3×3 の平均                                                          | 既定  |
|           | ヒストグラム、固定重みの値            | 6×6、1.0                                                             | 既定  |
| 信頼度に基づく重み | 基準の重みと式                  | 0.5、 $\max(0, 1 - \text{BC}_{\text{col}} / \text{BC}_{\text{tex}})$ | 既定  |
| 精緻化       | $\sigma_r$               | 0.12                                                                | 既定  |
|           | 色の補正                     | 32 段階、打ち切り±6、係数6 と3                                                 | 既定  |
|           | ガイドドフィルタ                 | 半径2、 $\epsilon = 10^{-3}$                                           | 既定  |
| 描画と評価     | 円板の半径、 $\epsilon$        | 12 画素 (幅640)、 $10^{-4}$                                             | 既定  |
| グラフカット    | 境界の重みと $\lambda$         | 式の $\omega_e$ 、 $\lambda = 32$                                      | 選択  |

## 参考文献

- [1] M. Potmesil and I. Chakravarty, “A lens and aperture camera model for synthetic image generation,” *Computer Graphics (Proc. SIGGRAPH ’81)*, vol. 15, no. 3, pp. 297–305, 1981.
- [2] N. Wadhwa, R. Garg, D. E. Jacobs, et al., “Synthetic depth-of-field with a single-camera mobile phone,” *ACM Trans. Graphics*, vol. 37, no. 4, Art. 64, 2018. doi:10.1145/3197517.3201329
- [3] Y. Y. Boykov and M.-P. Jolly, “Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images,” in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, vol. 1, pp. 105–112, 2001.
- [4] C. Rother, V. Kolmogorov, and A. Blake, “‘GrabCut’: Interactive foreground extraction using iterated graph cuts,” *ACM Trans. Graphics*, vol. 23, no. 3, pp. 309–314, 2004.
- [5] P. Salembier and L. Garrido, “Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval,” *IEEE Trans. Image Process.*, vol. 9, no. 4, pp. 561–576, 2000. doi:10.1109/83.841934
- [6] L. Guigues, J. P. Cocquerez, and H. Le Men, “Scale-sets image analysis,” *Int. J. Computer Vision*, vol. 68, no. 3, pp. 289–317, 2006.
- [7] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, “Contour detection and hierarchical image segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, 2011.
- [8] B. Ravi Kiran and J. Serra, “Global–local optimizations by hierarchical cuts and climbing energies,” *Pattern Recognition*, vol. 47, no. 1, pp. 12–24, 2014. doi:10.1016/j.patcog.2013.05.012
- [9] V. Lempitsky, A. Vedaldi, and A. Zisserman, “A pylon model for semantic segmentation,” in *Advances in Neural Information Processing Systems (NIPS)*, vol. 24, 2011.

- [10] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 147–159, 2004. doi:10.1109/TPAMI.2004.1262177
- [11] Y. Li, J. Sun, C.-K. Tang, and H.-Y. Shum, "Lazy snapping," *ACM Trans. Graphics*, vol. 23, no. 3, pp. 303–308, 2004. doi:10.1145/1015706.1015719
- [12] P. Kohli and P. H. S. Torr, "Dynamic graph cuts for efficient inference in Markov random fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 12, pp. 2079–2088, 2007. doi:10.1109/TPAMI.2007.1128
- [13] V. Lempitsky, P. Kohli, C. Rother, and T. Sharp, "Image segmentation with a bounding box prior," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2009. doi:10.1109/ICCV.2009.5459262
- [14] H. Yu, Y. Zhou, H. Qian, et al., "LooseCut: Interactive image segmentation with loosely bounded boxes," arXiv:1507.03060, 2015.
- [15] N. Xu, B. Price, S. Cohen, J. Yang, and T. Huang, "Deep GrabCut for object selection," in *Proc. British Machine Vision Conf. (BMVC)*, 2017. arXiv:1707.00243
- [16] J. C. Gower and G. J. S. Ross, "Minimum spanning trees and single linkage cluster analysis," *J. Royal Statistical Society, Series C (Applied Statistics)*, vol. 18, no. 1, pp. 54–64, 1969. doi:10.2307/2346439
- [17] J. Cousty, L. Najman, Y. Kenmochi, and S. Guimarães, "Hierarchical segmentations with graphs: Quasi-flat zones, minimum spanning trees, and saliency maps," *J. Mathematical Imaging and Vision*, vol. 60, no. 4, pp. 479–502, 2018. doi:10.1007/s10851-017-0768-7
- [18] L. Najman, J. Cousty, and B. Perret, "Playing with Kruskal: Algorithms for morphological trees in edge-weighted graphs," in *Mathematical Morphology and Its Applications to Signal and Image Processing (ISMM 2013)*, Lecture Notes in Computer Science, pp. 135–146, 2013. doi:10.1007/978-3-642-38294-9\_12
- [19] P. Soille, "Constrained connectivity for hierarchical image partitioning and simplification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 7, pp. 1132–1145, 2008. doi:10.1109/TPAMI.2007.70817
- [20] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Computer Vision*, vol. 59, no. 2, pp. 167–181, 2004. doi:10.1023/B:VISI.0000022288.19776.77
- [21] S. X. Hu, C. K. I. Williams, and S. Todorovic, "Tree-cut for probabilistic image segmentation," arXiv:1506.03852, 2015.
- [22] T. Porter and T. Duff, "Compositing digital images," *Computer Graphics (Proc. SIGGRAPH '84)*, vol. 18, no. 3, pp. 253–259, 1984.
- [23] A. Levin, D. Lischinski, and Y. Weiss, "A closed-form solution to natural image matting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 228–242, 2008.
- [24] C. Rhemann, C. Rother, J. Wang, M. Gelautz, P. Kohli, and P. Rott, "A perceptually motivated online benchmark for image matting," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 1826–1833, 2009. doi:10.1109/CVPR.2009.5206503
- [25] A. R. Smith and J. F. Blinn, "Blue screen matting," in *Proc. SIGGRAPH '96*, pp. 259–268, 1996.
- [26] E. Reinhard, M. Ashikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Computer Graphics and Applications*, vol. 21, no. 5, pp. 34–41, 2001. doi:10.1109/38.946629
- [27] D. L. Ruderman, T. W. Cronin, and C.-C. Chiao, "Statistics of cone responses to natural images: implications for visual coding," *J. Optical Society of America A*, vol. 15, no. 8, pp. 2036–2045, 1998.
- [28] C. Elkan and K. Noto, "Learning classifiers from only positive and unlabeled data," in *Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, pp. 213–220, 2008.
- [29] A. K. Menon, B. van Rooyen, C. S. Ong, and R. C. Williamson, "Learning from corrupted binary labels via class-probability estimation," in *Proc. Int. Conf. Machine Learning (ICML)*, PMLR vol. 37, pp. 125–134, 2015.
- [30] C. Scott, G. Blanchard, and G. Handy, "Classification with asymmetric label noise: consistency and maximal denoising," in *Proc. Annual Conf. Learning Theory (COLT)*, PMLR vol. 30, pp. 489–511, 2013.
- [31] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bulletin of the Calcutta Mathematical Society*, vol. 35, pp. 99–109, 1943.

- [32] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. Academic Press, 1990.
- [33] R. T. Collins, Y. Liu, and M. Leordeanu, "Online selection of discriminative tracking features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1631–1643, 2005. doi:10.1109/TPAMI.2005.205
- [34] A. Kirillov, E. Mintun, N. Ravi, et al., "Segment anything," in *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, pp. 3992–4003, 2023. doi:10.1109/ICCV51070.2023.00371
- [35] J. Li, J. Jain, and H. Shi, "Matting Anything," arXiv:2306.05399, 2023.
- [36] H. Knutsson and C.-F. Westin, "Normalized and differential convolution," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 515–523, 1993. doi:10.1109/CVPR.1993.341081
- [37] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object classes (VOC) challenge," *Int. J. Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010. doi:10.1007/s11263-009-0275-4
- [38] O. Boiman, E. Shechtman, and M. Irani, "In defense of nearest-neighbor based image classification," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2008. doi:10.1109/CVPR.2008.4587598
- [39] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.
- [40] B. Mahdian and S. Saic, "Using noise inconsistencies for blind image forensics," *Image and Vision Computing*, vol. 27, no. 10, pp. 1497–1503, 2009. doi:10.1016/j.imavis.2009.02.001
- [41] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," *ACM Trans. Graphics*, vol. 26, no. 3, Art. 96, 2007.
- [42] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, 2013. doi:10.1109/TPAMI.2012.213
- [43] H. Shimodaira, "Improving predictive inference under covariate shift by weighting the log-likelihood function," *Journal of Statistical Planning and Inference*, vol. 90, no. 2, pp. 227–244, 2000.
- [44] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, 2004. doi:10.1109/TPAMI.2004.60