2025 年度 修士論文

論文題目
大腸内視鏡動画セット C3VD と内視鏡フード情報を利用したポリープの大きさ推定
(Polyp Size Estimation Using the Colonoscopy 3D Video Dataset and Endoscopic Hood Geometry)

指導教員
舟橋　健司　准教授

名古屋工業大学大学院　工学研究科　情報工学系プログラム
2023 年度入学 35414034 番

菊地 遼

目 次

第1章　はじめに

厚生労働省の統計 [1] によれば，2024 年の日本における死亡者数は 160 万 5378 人であり，そのうち癌による死亡は全体の 23.9%を占める 38 万 4111 人である．さらに，癌は昭和 56 年以降，長年にわたり死因の第 1 位であり，社会的にも重大な課題である．癌は早期に発見し適切な治療を受けることで死亡リスクを大きく低減できることが知られている．例えば，大腸癌治療ガイドライン [2] では，固有筋層を超えて浸潤しているもののリンパ節転移・遠隔転移を認めない Stage II 以下の大腸癌に対し，治療後 5 年生存率が 84.8%であることが示されており，早期発見・早期治療の重要性が示されている．

大腸に生じるポリープは，非腫瘍性ポリープと腫瘍性ポリープに大別される．非腫瘍性ポリープには炎症性ポリープや過形成性ポリープなどがあり，正常細胞の増殖により隆起を形成するものの，癌化の可能性は一般に低い．これに対し腫瘍性ポリープには，癌化リスクを有する悪性ポリープと，癌化しにくい良性ポリープが含まれる．悪性が疑われる場合には摘出が必要となる一方，良性ポリープまで同様に摘出すると患者負担の増大につながる．そのため，癌の早期発見に加え，不要な処置を避けて患者の負担を減らすためにも，良悪性の判定誤りは可能な限り避ける必要がある．臨床現場では，内視鏡観察により形状・色調・大きさ等を総合してポリープの性状判断が行われるが，なかでも大きさは重要な指標の一つである．特にポリープ径が 10 mm 以上の場合，悪性の可能性が高まることが報告されている．しかし，内視鏡画像は二次元であり直接的な奥行き情報を持たない．さらに，内視鏡は病変発見を目的として広角に設計されるため，画像には広角レンズ特有の歪みも生じる．このような撮影条件下では，術者が経験に基づいておおよその大きさを見積もることはできても，ポリープの大きさや形状を定量化することは困難である．

ポリープの大きさを計測する試みとして，内視鏡からレーザ光を投射する方法 [3],

注射針に目盛を付与する方法 [4]，スネアに目盛を付与する方法 [5] などが提案されてきた．しかし，これらは専用機器や専用内視鏡の導入を要する場合が多く，汎用的な内視鏡検査へ広く適用するうえでの制約となる．そのため，一般的な内視鏡画像から大きさ・形状を推定できる方法が求められている．この課題に対し，内視鏡画像からポリープの三次元形状を復元し，そこから大きさや形状を推定する研究が進められてきた．三次元形状復元の代表的手法の一つとして，画像の陰影から形状を復元する Shape-from-Shading（SFS）[6] がある．SFS は，反射特性と照明条件が既知であるという仮定のもと，輝度と形状の関係を記述する光学式を三次元形状に関する偏微分方程式として定式化し，これを解くことで形状を推定する．Verbeek らは，視線方向と平行な照明下における Lambert 反射を仮定すると，SFS 問題が Eikonal 方程式として扱えることを示し [7]，同方程式を解くことで三次元形状復元が可能であることを示した．ただし Eikonal 方程式の解法として反復計算を用いると計算負荷が大きい．そこで Sethian らは高速解法として Fast Marching Method（FMM）[8] を提案し，視点に近い画素から遠い方向へ逐次的に解を更新することで，高速に形状復元を行う枠組みを与えた．さらに Kimmel らは陰影情報から Eikonal 方程式を導き，FMM で解くことにより高速な三次元復元を実現する手法を提案した [9]．その後，透視投影条件へ拡張するための更新式が提案され [10][11]，内視鏡画像のような透視投影環境への適用可能性が広がった．一方で，単一画像に基づく形状復元には絶対スケールが定まらないという課題がある．すなわち，画像中に大きさの基準となる参照物が存在しない場合，復元形状は相対的な形状に留まり，実寸での大きさ推定が困難となる．内視鏡画像への応用研究としては，点光源条件を考慮した拡張 [12]，光学的制約と幾何学的制約を併用した最適化による高精度化 [13]，複数光源への拡張 [14] などが提案されてきた．しかし，最適化を含む手法は処理時間が増大しやすいという問題を抱える．これに対し，光学的制約と幾何学的制約が同一の高さを与えるという関係を利用して Eikonal 方程式に基づく式を導出し，高精度かつ高速な復元を狙う手法も提案されている [15]．また，ポリープの大きさを推定する手法として，内視鏡画像に比較的高頻度で含まれる血管を参照物とすることで大きさを推定する手法も提案されている [16]．単一画像からカメラに対して水平な血管領域を検出し，血管幅を既知と仮定して深度と反射係数を推定し，周辺肉壁へ拡張

する手法も提案されている [17]．ただしこの方法は，水平な血管領域が画像内に存在することを前提としている．この制約を緩和するため，血管領域の傾きをアフィン変換で補正し，仮定した血管幅に合う反射係数を最適化で求める方法 [18] も報告されているが，血管が近傍に存在しない場合には適用できないことや，参照とする血管の選択や幅の仮定に起因して誤差が生じうることが課題として挙げられる．

これまでに筆者らは，FMM に基づく形状推定手法において，内視鏡先端に装着されるフードの幾何学情報を活用することで反射係数を推定する方法を提案した [19]．フードは装着状態で画像内のほぼ一定位置に写り込み，レンズとフードの相対配置が一定となるため，画像内の参照物として利用できる．この性質を用い，反射率を一様化した Lambert 画像上でフード領域の輝度と三次元情報を参照することで反射係数を推定し，推定した反射係数に基づいてポリープの大きさ推定および三次元形状復元を行う枠組みを構築した．しかし，実画像に対する評価を進めると，推定結果にモデル化誤差や撮影条件差に起因すると考えられる過小推定の傾向を含む誤差を確認した．

本研究では，単一の内視鏡画像からポリープの大きさを推定するという課題において，FMM に基づいた形状推定枠組みとは別に，機械学習ベースの深度推定枠組みを提案する．C3VD データセットを教師データとして単眼深度推定モデルを学習し，得られた深度に対してフードの既知情報を用いてスケール情報を与えることで，通常内視鏡画像から mm 単位の深度推定を行う．さらに，FMM ベースおよび機械学習ベースのいずれにも適用可能なニューラルネットワークによる誤差補正を導入し，ポリープ大きさ推定の精度向上を図る．

第2章　背景知識

本章では，大腸内視鏡画像におけるポリープの大きさ・形状推定の課題および，単眼深度推定に内在するスケール不定性について述べる．さらに，先行研究で提案した推定手法 [19] について述べる．

2.1　現状の課題

ポリープの大きさ・形状を画像から定量化するには，深度を含む三次元情報が必要である．しかし内視鏡では，照明条件や反射の影響により輝度変化が形状変化と対応しにくく，加えて撮影距離の変動により，単眼画像から絶対スケールを一意に定めることが難しい．以下では，まず内視鏡環境における大きさ・形状推定の困難性を述べ，続いて単眼深度推定のスケール不定性について説明する．

2.1.1　内視鏡における大きさ・形状推定

大腸内視鏡検査において，ポリープの大きさや形状は診断や治療方針の決定に関わる重要な情報である．しかし通常の内視鏡で得られる情報は二次元画像が主であり，対象までの深度や三次元形状を直接観測できない．このため，画像から mm 単位の定量的な大きさや三次元形状を推定することは本質的に難しい．内視鏡環境では距離や照射条件が明暗に強く影響し，さらに腸管内壁の湿潤さに起因する鏡面反射も頻繁に生じる．その結果，輝度変化が必ずしも形状変化のみを反映しないため，陰影から形状を復元する枠組みでは誤差が生じやすいという課題がある．

2.1.2　単眼深度推定とスケール不定性

近年，深層学習を用いた単眼深度推定が広く研究され，内視鏡画像への応用も進められている．しかし単眼画像からの深度推定には，透視投影に起因するスケール不定

性が存在する．すなわち，三次元空間内の点 (X, Y, Z) を一様に α 倍して $(\alpha X, \alpha Y, \alpha Z)$ としても，画像上の投影は変化しないため，単一画像のみから深度の絶対スケールを一意に決定できない．この性質は深層学習モデルにおいても同様であり，推定深度 $\hat{Z}$ が真の深度 Z に対して $\hat{Z} \approx \alpha Z$ のように未知のスケール α を含む相対深度を推定する手法が多い．したがって，単眼深度推定を mm 単位の大きさ推定へ活用するためには，スケール α を決める追加情報が必要となる．ただし内視鏡では，非剛体性や特徴点不足，撮影距離・角度の変動，反射の影響などにより，スケールの特定が困難である．そのため，内視鏡という撮影環境に適した参照情報を与え，スケールの安定化を図ることが重要となる．

2.2　先行研究：FMM に基づいた大きさ・形状推定

内視鏡画像のスケール決定に関しては，鉗子や血管等の既知物体の写り込みを利用する方法が知られているが，臨床運用上これらが常に画面内に存在するとは限らない．そこで，先行研究では内視鏡フードという内視鏡先端に装着されるアタッチメントに着目した [19]．内視鏡フードは，図 2.1，2.2 に示すように内視鏡先端に取り付ける補助デバイスである．消化管内を内視鏡で観察する際，粘液や血液，脂肪などがレンズに付着して視野が低下し，観察が困難となることがある．この汚れを除去するには，検査を一時中断して内視鏡を抜去し，洗浄後に再挿入する必要があり，医師の負担増大や患者の侵襲につながる．内視鏡フードは，内視鏡の周囲を囲いレンズの汚れや曇りを抑制することで，これらの問題を軽減し観察性を向上させる目的で用いられる．フードの内径と突出長を参照情報として利用することで，高価な追加器具に依存せずにスケールを導入する推定手法を提案した．この手法では，画像中に写るフードの輝度情報から反射係数を推定することで，FMM に基づいて形状復元したポリープのスケールを定め，大きさを推定した．処理のフローチャートを図 2.3 に示す．

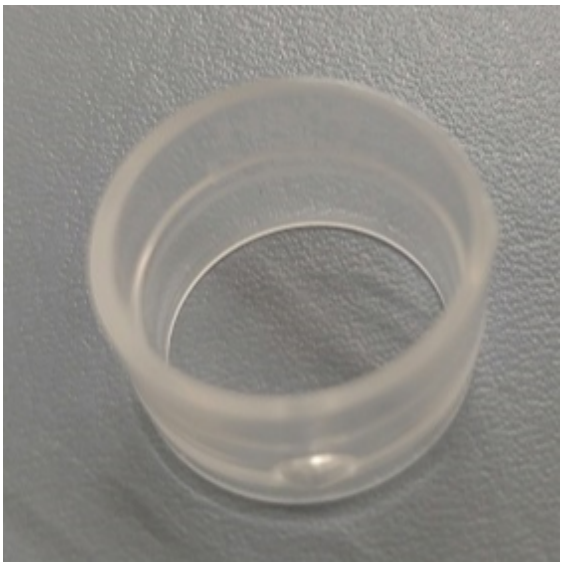


図 2.1: 内視鏡フード

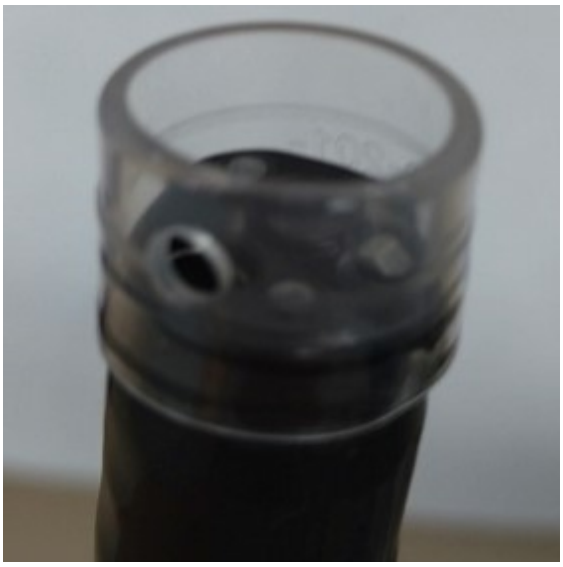


図 2.2: フードを取り付けた内視鏡

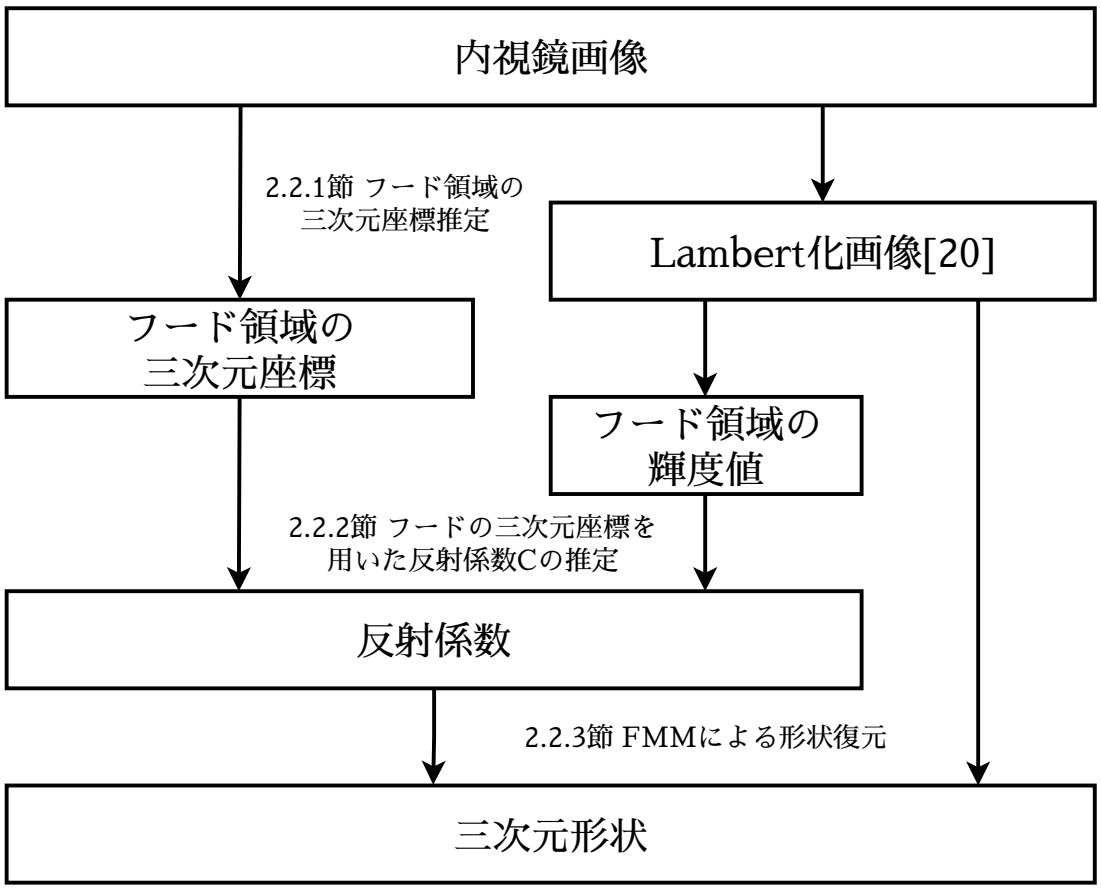


図 2.3: 従来手法の処理手順

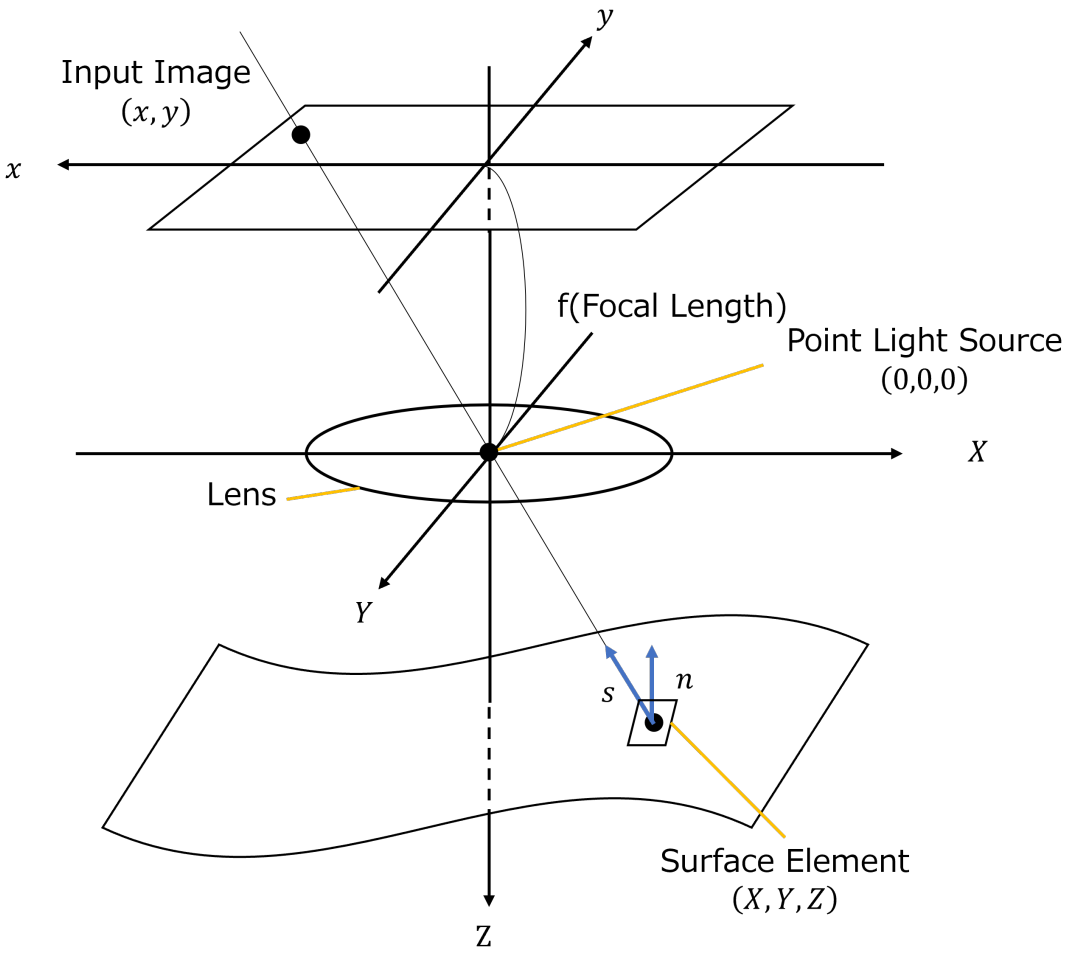


図 2.4: 透視投影モデル

2.2.1　フード領域の三次元座標推定

本節では，フードの既知寸法を参照情報として用いるため，画像中に写り込むフード領域の画素座標からフード表面の三次元座標を推定する．その際，カメラモデルとして透視投影を仮定し，三次元座標 (X, Y, Z) と画像座標 (x, y) の対応を述べる．

透視投影とは，三次元物体を平面上で表現するための投影法の一つである．透視投影では，視点から物体までの距離が長い場合は物体が小さく，距離が短い場合には物体が大きく投影され，対象物を目で見た時に近い表現が得られる．透視投影のモデルを図 2.4 に示す．透視投影における二次元画像の x, y 座標は，式 (2.1) によっ

て求められる．f は焦点距離，X,Y,Z は物体の三次元座標である．

$$\begin{aligned} x &= \frac{X}{Z} f \\ y &= \frac{Y}{Z} f \end{aligned} \tag{2.1}$$

フードの形状および突出部の長さは定まっているため，それらの情報に加えて，内視鏡によって撮影された二次元画像のフード部分の各座標 $x_{\mathrm{h}}, y_{\mathrm{h}}$ を用いることにより，フード部分の各画素における三次元座標 $X_{\mathrm{h}}, Y_{\mathrm{h}}, Z_{\mathrm{h}}$ を推定できる．

フードの内側の半径を R，焦点距離を f とすると，透視投影の環境下で次式が成り立つ．

$$\begin{cases} X_{\mathrm{h}}^2 + Y_{\mathrm{h}}^2 = R^2 \\ X = \dfrac{x_{\mathrm{h}}}{f} Z_{\mathrm{h}}, \quad Y = \dfrac{y_{\mathrm{h}}}{f} Z_{\mathrm{h}} \end{cases} \tag{2.2}$$

式 (2.2) を整理し，フードの奥行き Z について解くと次式が得られる．

$$Z = \frac{fR}{\sqrt{x^2 + y^2}} \tag{2.3}$$

なお，奥行き Z が負となる解は画像内に映ることがないため，正の解を採用する．式 (2.2)，(2.3) をまとめると，フードの三次元座標は $(X_{\mathrm{h}}, Y_{\mathrm{h}}, Z_{\mathrm{h}})$ を用いて以下のように求められる．

$$(X_{\mathrm{h}}, Y_{\mathrm{h}}, Z_{\mathrm{h}}) = \left(\frac{x_{\mathrm{h}}}{f} Z_{\mathrm{h}},\ \frac{y_{\mathrm{h}}}{f} Z_{\mathrm{h}},\ \frac{fR}{\sqrt{x_{\mathrm{h}}^2 + y_{\mathrm{h}}^2}} \right) \tag{2.4}$$

2.2.2　フードの三次元座標を用いた反射係数 C の推定

次に，フード部分の三次元座標を用いた反射係数の推定方法について説明する．反射係数は，フード領域の画素を対象として最小二乗推定する．手法 [20] で得られる反射率一様の Lambert 化画像について，フード領域の輝度と理想輝度の差の二乗和が最小となる反射係数を求める．点光源，透視投影，反射係数一様の条件下において，フード部分の各点における法線ベクトル $\boldsymbol{n}$，光源方向ベクトル $\boldsymbol{s}$ およびその内

積は以下のように得られる．

$$\begin{cases} \boldsymbol{n} = \dfrac{(-X_{\mathrm{h}}, -Y_{\mathrm{h}}, 0)}{R} \\ \boldsymbol{s} = \dfrac{(-X_{\mathrm{h}}, -Y_{\mathrm{h}}, -Z_{\mathrm{h}})}{\sqrt{X_{\mathrm{h}}^2 + Y_{\mathrm{h}}^2 + Z_{\mathrm{h}}^2}} \\ \boldsymbol{s} \cdot \boldsymbol{n} = \dfrac{X_{\mathrm{h}}^2 + Y_{\mathrm{h}}^2}{R\sqrt{X_{\mathrm{h}}^2 + Y_{\mathrm{h}}^2 + Z_{\mathrm{h}}^2}} \end{cases} \tag{2.5}$$

なお，フードまでの距離は $\sqrt{X_{\mathrm{h}}^2 + Y_{\mathrm{h}}^2 + Z_{\mathrm{h}}^2}$ であるため，これらのことから理想的な輝度値 E は次式によって求めることができる．なお，R はフードの内側の半径，C は反射係数を表す．

$$\begin{aligned} E &= C\frac{(\boldsymbol{s} \cdot \boldsymbol{n})}{X_{\mathrm{h}}^2 + Y_{\mathrm{h}}^2 + Z_{\mathrm{h}}^2} \\ &= C\frac{X_{\mathrm{h}}^2 + Y_{\mathrm{h}}^2}{R(X_{\mathrm{h}}^2 + Y_{\mathrm{h}}^2 + Z_{\mathrm{h}}^2)^{\frac{3}{2}}} \end{aligned} \tag{2.6}$$

また，式 (2.6) を式 (2.2) によって簡単にすると次式が得られる．

$$E = C\frac{R}{(R^2 + Z_{\mathrm{h}}^2)^{\frac{3}{2}}} \tag{2.7}$$

ここで，入力された Lambert 化画像から得られる，フード部分の画素 (i, j) における輝度値を E_{ij} と表す．また，画素 (i, j) における理想的な輝度値 A_{ij} を式 (2.7) から求めると，次式が得られる．なお，簡単のため，ϕ_{ij} を次式とする．

$$\phi_{ij} = \frac{R}{(R^2 + Z_h(i, j)^2)^{\frac{3}{2}}} \tag{2.8}$$

$$A_{ij} = C\phi_{ij} \tag{2.9}$$

このとき，各画素における E_{ij} と A_{ij} の二乗誤差和 ϵ は次式のようになる．

$$\epsilon = \sum_{(i,j) \in H} (E_{ij} - C\phi_{ij})^2 \tag{2.10}$$

ここで H はフード領域の画素集合である．

次に ϵ が最小となる反射係数 C を求める．式 (2.10) を変形すると以下のように

なる．

$$\begin{aligned}\epsilon &= \sum_{(i,j)\in H} \{E_{ij} - C\phi_{ij}\}^2 \\ &= \sum E_{ij}^2 - 2C\sum E_{ij}\phi_{ij} + C^2\sum \phi_{ij}^2 \\ &= \sum \phi_{ij}^2 \left\{ C^2 - 2C\frac{\sum E_{ij}\phi_{ij}}{\sum \phi_{ij}^2} + \left(\frac{\sum E_{ij}\phi_{ij}}{\sum \phi_{ij}^2}\right)^2 \right\} + \left\{ \sum E_{ij}^2 - \frac{\left(\sum E_{ij}\phi_{ij}\right)^2}{\sum \phi_{ij}^2} \right\} \\ &= \sum \phi_{ij}^2 \left(C - \frac{\sum E_{ij}\phi_{ij}}{\sum \phi_{ij}^2} \right)^2 + \left\{ \sum E_{ij}^2 - \frac{\left(\sum E_{ij}\phi_{ij}\right)^2}{\sum \phi_{ij}^2} \right\}\end{aligned} \tag{2.11}$$

式 (2.11) において反射係数 C に関わる項を抜き出すと次式が得られる．

$$\epsilon_2 = \sum_{(i,j)\in H} \phi_{ij}^2 \left(C - \frac{\sum_{(i,j)\in H} E_{ij}\phi_{ij}}{\sum_{(i,j)\in H} \phi_{ij}^2} \right)^2 \tag{2.12}$$

よって，二乗誤差 ϵ を最小にする式 C は次式によって得られる．

$$C = \frac{\sum_{(i,j)\in H} E_{ij}\phi_{ij}}{\sum_{(i,j)\in H} \phi_{ij}^2} \tag{2.13}$$

2.2.3　FMM（Fast Marching Method）による形状復元

FMM における三次元形状復元は，既知点の表面勾配パラメータ p, q および奥行き Z に基づき，隣接点の p, q, Z を逐次的に推定する．未知点がなくなるまで更新を繰り返すことで，三次元形状を復元する．図 2.5 に簡単な処理の過程を示す．簡単のため，

$$V = \frac{f^2}{(x^2 + y^2 + f^2)^{\frac{3}{2}}} \tag{2.14}$$

としたとき，奥行き Z を求める式は以下のように表される．

$$\begin{aligned} & E = \frac{CV(-px - qy + f)}{Z^2(p^2 + q^2 + 1)^{\frac{1}{2}}} \\ \Leftrightarrow \quad & Z^2 = \frac{CV(-px - qy + f)}{E(p^2 + q^2 + 1)^{\frac{1}{2}}} \\ \Rightarrow \quad & Z = \sqrt{\frac{CV(-px - qy + f)}{E(p^2 + q^2 + 1)^{\frac{1}{2}}}} \end{aligned} \tag{2.15}$$

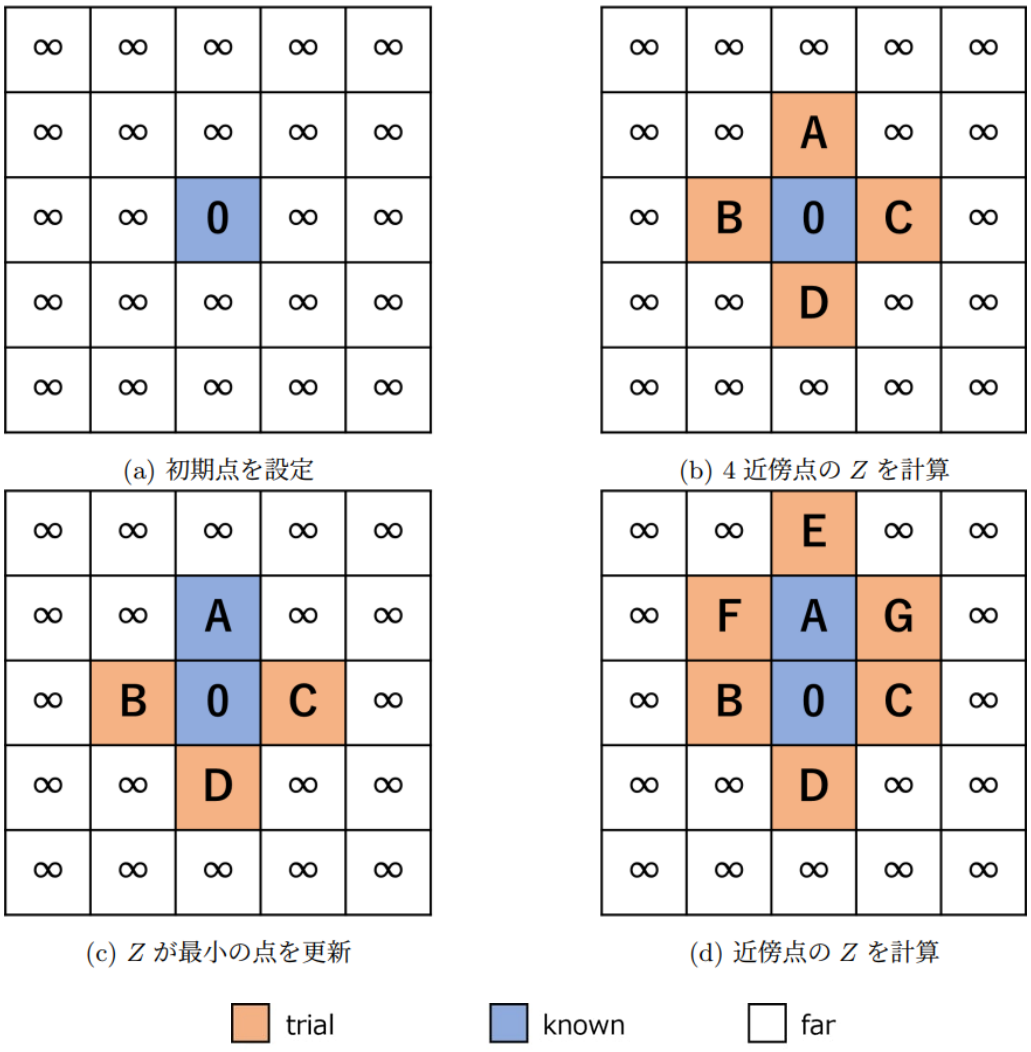


図 2.5: FMM の流れ (文献 [13])

また，FMM の更新式を導出するために，次式で定義される量 A を用いる．

$$A = \frac{V^2(f - p_k x_t - q_k y_t)^6}{Z_k^4(f - p_k x_k - q_k y_k)^4} \tag{2.16}$$

ここで，k は既知点，t は k に隣接する未知点（trial）を表す．このとき，$p \approx p_k,\ q \approx q_k$ という近似を用いて変形を行うことで，Eikonal 方程式の形として次式が得られる．

$$\sqrt{p^2 + q^2} = \sqrt{\left(\frac{C}{E}\right)^2 A - 1} \tag{2.17}$$

次に，式 (2.17) を Z の x, y 方向の差分式に置き換える．i, j は注目点の画素座標として，画素座標 (i, j) における式 (2.17) の右辺を F_{ij} とする．

$$F_{ij} = \sqrt{\left(\frac{C}{E_{ij}}\right)^2 A_{ij} - 1} \tag{2.18}$$

$$\sqrt{\max(p_-, -p_+, 0)^2 + \max(q_-, -q_+, 0)^2} = F_{ij} \tag{2.19}$$

$$\begin{cases} p_- = \dfrac{Z_{i,j} - Z_{i-1,j}}{X_{i,j} - X_{i-1,j}} \\ p_+ = \dfrac{Z_{i+1,j} - Z_{i,j}}{X_{i+1,j} - X_{i,j}} \\ q_- = \dfrac{Z_{i,j} - Z_{i,j-1}}{Y_{i,j} - Y_{i,j-1}} \\ q_+ = \dfrac{Z_{i,j+1} - Z_{i,j}}{Y_{i,j+1} - Y_{i,j}} \end{cases} \tag{2.20}$$

この時，式 (2.19) を奥行き Z について解くことにより，次式が得られる．

$$Z_{ij} = \begin{cases} \dfrac{Z_a + Z_b + \sqrt{2F_{ij}^2 - (Z_a - Z_b)^2}}{2} & (|Z_a - Z_b| < F_{ij}) \\ \min(Z_a, Z_b) + F_{ij} & (|Z_a - Z_b| \geq F_{ij}) \end{cases} \tag{2.21}$$

$$Z_a = \min(Z_{i-1,j}, Z_{i+1,j}), \qquad Z_b = \min(Z_{i,j-1}, Z_{i,j+1}) \tag{2.22}$$

また，FMM によって三次元形状復元を行う際，p, q, Z が既知である初期点が必要である．点光源・透視投影の環境下において，輝度が極大となる点では光源方向ベクトル $\boldsymbol{s}$ が法線方向ベクトル $\boldsymbol{n}$ に一致することが示されているため，輝度極大点では p, q は次式のように表せる．

$$p = -\frac{x}{f}, \qquad q = -\frac{y}{f} \tag{2.23}$$

また，奥行き Z は式 (2.23) を光学的制約式 (2.15) に代入することで求めることができるため，輝度極大点を初期点として，以下の FMM のアルゴリズムに基づく処理により復元を行う．

- Step 1. 輝度極大点を初期点とし，表面勾配パラメータ p, q および，奥行き Z を求める．
- Step 2. 初期点以外の奥行き Z を $Z = \infty$ とする．
- Step 3. 既知の点の 4 近傍かつ $Z = \infty$ の点における仮の奥行き Z を更新式 (2.21) によって求める．
- Step 4. Step3 にて求めた Z を用いて x, y 方向の差分式 (2.20) から，4 近傍の (p, q) を更新する．
- Step 5. Step4 にて求めた p, q を用いて式 (2.21) から，4 近傍の Z を更新する．
- Step 6. Step5 にて求めた Z が最小である画素を既知とする．
- Step 7. 全ての画素の Z が計算されるまで Step3〜Step6 を繰り返す．

第3章　提案手法

本章では，まず単眼深度推定モデルを学習し，得られた推定深度に対してフードの既知情報を基準としたスケールを導入することで，通常内視鏡画像から絶対深度推定を行う方法を述べる．次に，従来手法および提案手法で得られるポリープ大きさの推定結果に対する，ニューラルネットワークによる大きさ推定誤差の補正手法を述べる．

先行研究では，レンズに対するフードの三次元情報を用いて反射係数を求めることでスケールを推定し，FMM によりポリープの大きさを推定した．一方，本章で述べる機械学習による深度推定では，推定される深度のスケールが内視鏡の環境の差により異なるため，フードの既知情報から焦点距離を得ることでスケールを付与した．

3.1　C3VD を用いた単眼深度推定に基づく大きさ推定

本節では，C3VD データセットを教師データとして単眼深度推定モデルを学習し，通常内視鏡画像から絶対深度を推定する，機械学習ベースの枠組みについて述べる．処理のフローチャートを図 3.1 に示す．

3.1.1　Colonoscopy 3D Video Dataset

本研究では，深層学習による単眼深度推定の教師データとして Colonoscopy 3D Video Dataset (以下，C3VD) を使用する [21],[22]．C3VD は，シリコン製の大腸模型を臨床用高精細内視鏡で撮影した動画に対し，画素単位の教師情報（深度，法線，オクルージョン，光フロー，カメラ姿勢など）を付与したデータセットである．本研究では，RGB 画像と対応する深度画像を用いて，教師あり学習による絶対深度推定を行う．

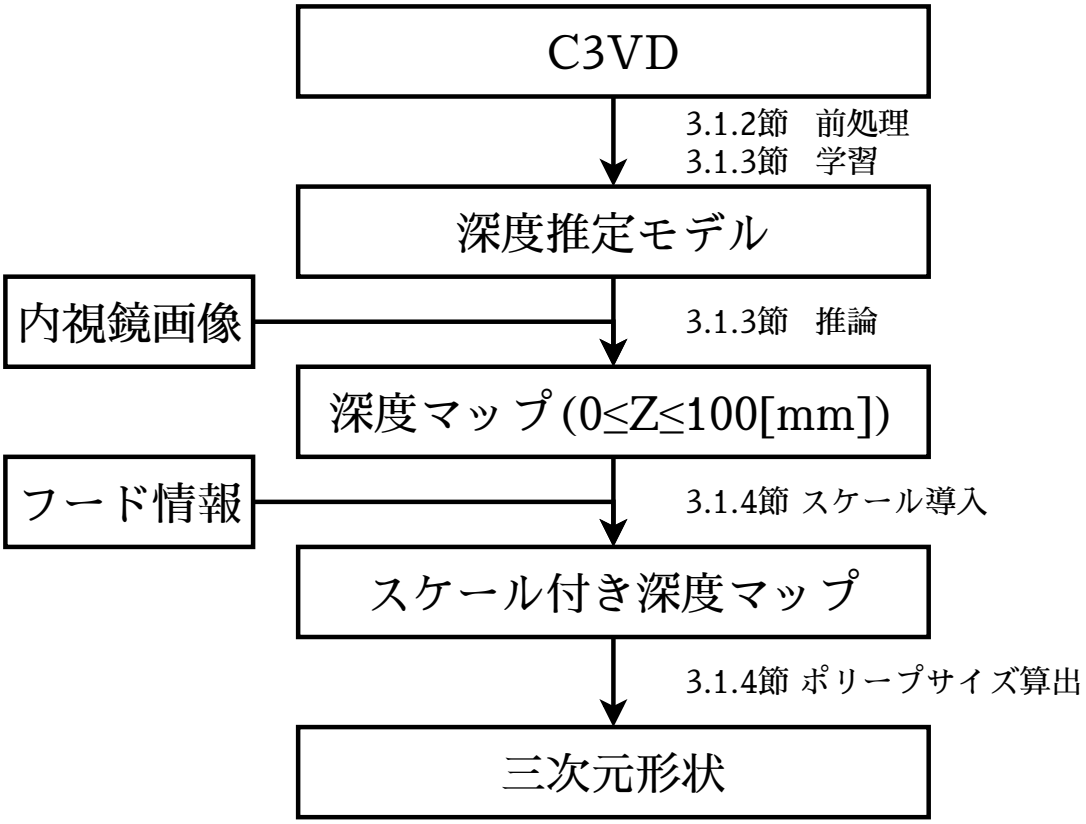


図 3.1: 提案手法の処理手順

3.1.2　前処理

C3VD が提供する教師深度は 16-bit 画像として与えられ，画素値 $I_{ij} \in [0, 65535]$ が深度 Z_{ij} へ線形に対応する．C3VD では，深度の上限を 100 mm として深度画像を表現しているため，次式により Z_{ij} [mm] へ復号する．

$$Z_{ij} = 100 \cdot \frac{I_{ij}}{65535} \tag{3.1}$$

なお，$I_{ij} = 65535$ に飽和した画素は，真値が 100 mm より遠方である可能性があるため，損失計算から除外する．

RGB 画像はリサイズ後にテンソル化し，ImageNet の平均・標準偏差で正規化する．教師深度は式 (3.1) で復号した後，RGB と同一解像度へ最近傍補間でリサイズし，学習時の入力と教師の画素対応を揃える．また，動画の特性上連続したフレームは似通った画像であることが多いため，動画フレームを 1/10 サンプリングで間引

いて学習に用いる．

3.1.3　学習と推論

深度推定には，ResNet34 をエンコーダとする U-Net 型ネットワーク（ResNet-UNet）を用いる．また，入力 RGB 画像に対し，画素ごとの推定深度 $\hat{Z}$ [mm] を出力する．教師深度 Z と推定深度 $\hat{Z}$ の誤差として，L_1 損失を用いる．

$$\mathcal{L}_{\text{depth}} = \frac{1}{|\Omega|} \sum_{(i,j)\in\Omega} \left| Z_{ij} - \hat{Z}_{ij} \right| \tag{3.2}$$

ここで Ω は損失計算対象の画素集合である．

推論時は入力 RGB 画像を，単眼深度推定モデルにより深度マップ $\hat{Z}$ [mm] を得る．なお，本研究で用いる単眼深度推定モデルは，C3VD データセットに含まれる実模型画像を用いて学習されている．C3VD では内視鏡の焦点距離 f_{C3VD} [px] が既知であり，モデルはこの焦点距離を前提とした透視投影幾何に基づいて深度を学習している．

3.1.4　推定深度を用いたポリープサイズ算出

C3VD と実画像では，使用する内視鏡の光学系が異なるため，焦点距離が一致しない．透視投影モデルにおいては，焦点距離と深度はスケール的に結合しており，学習時と推論時で焦点距離が異なる場合，推定される深度には焦点距離の比に起因するスケール差が生じる．そこで，本研究では C3VD に基づいて推定された深度 $\hat{Z}$ [mm] に対し，実画像の光学系と整合するようなスケールを導入する．実画像の焦点距離を f_{real} [px] とすると，補正後の深度 Z_{est} [mm] は次式で与えられる．

$$Z_{\text{est}} = \hat{Z} \frac{f_{\text{real}}}{f_{\text{C3VD}}} \tag{3.3}$$

補正後の推定深度と画像上のピクセル長から，透視投影モデルに基づく相似関係を用いてポリープの大きさを算出する．ポリープの画像上でのピクセル径を S_{pixel} [px]，ポリープまでの深度を Z [mm] とする．このとき，有効焦点距離 f_{real} [px] が得られていれば，ポリープの大きさ S_{est} [mm] は次式で与えられる．

$$S_{\text{est}} = \frac{S_{\text{pixel}}\, Z}{f_{\text{real}}} \tag{3.4}$$

式 (3.3) および式 (3.4) より，最終的なポリープサイズ推定は，C3VD の焦点距離を前提として学習された深度の幾何関係を保持したまま実画像に適用されていることが分かる．本研究では，処理の意味を明確にするため深度のスケール特定とサイズ算出を段階的に記述している．式 (3.4) に必要な f_{real} は内視鏡フードの既知情報を用いて推定する．フードの内径を $D_{\mathrm{hood}}\,[\mathrm{mm}]$，突出長を $L_{\mathrm{hood}}\,[\mathrm{mm}]$，画像上で観測されるフード内径のピクセル長を $D_{\mathrm{pixel}}\,[\mathrm{px}]$ とする．透視投影モデルに基づく相似関係より，有効焦点距離 f_{real} は次式で与えられる．

$$f_{\mathrm{real}} = \frac{D_{\mathrm{pixel}}\, L_{\mathrm{hood}}}{D_{\mathrm{hood}}} \tag{3.5}$$

3.2　ニューラルネットワークによる大きさ推定の補正

従来手法および提案手法では，推定に用いる観測条件やモデル化誤差の影響により，ポリープ大きさ推定に誤差が生じうる．そこで本研究では，両手法で得られる推定結果を入力として，レーザ内視鏡で計測した正解大きさを回帰するニューラルネットワークを導入し，大きさ推定の誤差補正を行う．

3.2.1　学習の枠組み

モデルに基づいて得られる推定値 S_{est} を入力とし，レーザ内視鏡により計測した正解大きさ S_{gt} を目的変数とする．ニューラルネットワークを $g(\cdot)$ とすると，補正後推定値 $\hat{S}$ は次式で与えられる．

$$\hat{S} = g(\boldsymbol{x}) \tag{3.6}$$

ここで入力特徴量 $\boldsymbol{x}$ は，適用対象に応じて次の 2 通りを用いる．従来手法に対しては $\boldsymbol{x}_{\mathrm{conv}}$ を，提案手法に対しては $\boldsymbol{x}_{\mathrm{prop}}$ を入力とする．

$$\boldsymbol{x}_{\mathrm{conv}} = \begin{bmatrix} S_{\mathrm{est}} \\ C_{\mathrm{est}} \end{bmatrix} \tag{3.7}$$

$$\boldsymbol{x}_{\mathrm{prop}} = \begin{bmatrix} S_{\mathrm{est}} \\ Z_{\mathrm{est}} \\ S_{\mathrm{pixel}} \end{bmatrix} \tag{3.8}$$

3.2.2　ネットワーク構造

補正は少数の説明変数に基づく回帰問題であるため，多層パーセプトロン（MLP）により $g(\cdot)$ を構成する．入力層は二次元（S_{est}, C_{est}）または三次元（S_{est}, Z_{est}, S_{pixel}）とし，全結合層と ReLU を用いた 2 段の隠れ層を持つ構造とする．16 ユニットの全結合層，8 ユニットの全結合層を順に用い，最終層でスカラー出力 $\hat{S}$ を得る．

3.2.3　学習と推論

学習データは，各サンプルについて入力 $\boldsymbol{x}$ と正解 S_{gt} の組から構成する．損失関数は回帰誤差を最小化するため，平均二乗誤差（MSE）を用いる．

$$\mathcal{L}_{\mathrm{nn}} = \frac{1}{N}\sum_{n=1}^{N}\left(S_{\mathrm{gt}}^{(n)} - \hat{S}^{(n)}\right)^2 \tag{3.9}$$

ここで，N は学習に用いるサンプル数である．最適化には Adam を用い，学習率は 10^{-3}，最大エポック数は 500 とした．

また，入力特徴量は値域が大きく異なるため，学習の安定化を目的として標準化を行う．学習集合から特徴量ごとの平均 $\boldsymbol{\mu}$ と標準偏差 $\boldsymbol{\sigma}$ を算出し，

$$\tilde{\boldsymbol{x}} = \frac{\boldsymbol{x} - \boldsymbol{\mu}}{\boldsymbol{\sigma}} \tag{3.10}$$

としてスケーリングした入力 $\tilde{\boldsymbol{x}}$ をネットワークに与える．推論時も学習時と同一の $\boldsymbol{\mu}, \boldsymbol{\sigma}$ を用いて入力を標準化し，学習済みネットワーク $g(\cdot)$ により補正後推定値 $\hat{S}$ を得る．

第4章　評価

本章では初めに，第3章で提案したC3VDを教師データとする単眼深度推定手法の評価を行う．まず，C3VDの真値深度との比較により推定深度の妥当性を確認する．次に，得られた推定深度を用いてポリープの大きさを算出し，従来手法と提案手法を比較する．続いて，従来手法と提案手法の各手法に対してNN補正の有無を考慮した計4条件で評価を行う．

4.1　深度推定結果

本節では，第3章で学習した単眼深度推定モデルについて，C3VDの真値深度と比較し推定精度を確認する．C3VDはRGB画像に加えて真値深度が提供されるため，単眼深度推定に対して画素単位での定量的な比較が可能である．

図4.1に推定例を示す．推定深度は真値深度の大域的な分布を概ね再現しており，奥行き方向の勾配や形状の傾向が一致することが確認できる．一方で，境界近傍やテクスチャが乏しい領域では推定が平滑化され，局所的な誤差が残る傾向が見られた．これは単眼画像からの推定が局所的に不良条件（低コントラスト，反射，陰影の不安定さ）となる場合に，周辺情報に引きずられやすいことが一因と考えられる．

次に，図4.2に学習時の損失推移を示す．エポックの進行に伴い緩やかに収束し，0.7mm程度の誤差で安定していることを確認した．このことから，少なくともC3VD上では深度の回帰が破綻しておらず，真値深度との対応関係を学習できていると判断できる．

また，学習済みモデルを実画像に適用し，推定深度の定性的妥当性を確認した．実画像に対して学習済みモデルを適用する際，C3VDと実機でカメラ内部パラメータが異なるため，実際の内視鏡環境に適したスケールを導入する必要がある．ここでは，フード幾何と透視投影モデルに基づき実機の焦点距離を推定し，焦点距離比に

より内視鏡環境下に適したスケールを導入することで，正しい深度を推定する．

本研究では，フード内径 $D_{\mathrm{hood}} = 15\,\mathrm{mm}$，フード突出長 $L_{\mathrm{hood}} = 4\,\mathrm{mm}$ を既知とし，フード先端面までの距離を $Z = L_{\mathrm{hood}}$ とみなす．実画像においてフード内径が $D_{\mathrm{pixel}} = 1080\,\mathrm{px}$ であったことから，透視投影式 (3.5) より実機の焦点距離は以下のように算出できる．

$$f_{\mathrm{real}} = \frac{D_{\mathrm{pixel}}\, L_{\mathrm{hood}}}{D_{\mathrm{hood}}} = \frac{1080 \times 4}{15} \approx 288\ \mathrm{px} \tag{4.1}$$

一方，C3VD における焦点距離は $f_{\mathrm{C3VD}} = 769.24\,\mathrm{px}$ である．以上より，深度スケールが焦点距離に比例して変化すると仮定し，焦点距離比 α を用いて以下のようにスケールを導入した．

$$\alpha = \frac{f_{\mathrm{real}}}{f_{\mathrm{C3VD}}} \approx \frac{288}{769.24} \approx 0.3745 \tag{4.2}$$

$$Z_{\mathrm{est}} = \alpha\, \hat{Z} \tag{4.3}$$

実画像に対する深度推定例を図 4.3 に示す．

4.2　ポリープの大きさ推定結果

本節では，表 4.1 に示す 4 条件でポリープの大きさ推定精度を比較する．ポリープ大きさ推定の評価には愛知医科大学より提供された，レーザ内視鏡にて撮影された実画像 11 例を用いる．ポリープ大きさ推定の評価指標として，真値 S_{gt} と推定値 S_{est} の差に基づく MAE [mm] と MAPE [%] を用いる．

$$\mathrm{MAE} = \frac{1}{N} \sum_{n=1}^{N} |S_{\mathrm{gt}}^{(n)} - S_{\mathrm{est}}^{(n)}| \tag{4.4}$$

$$\mathrm{MAPE} = \frac{100}{N} \sum_{n=1}^{N} \frac{|S_{\mathrm{gt}}^{(n)} - S_{\mathrm{est}}^{(n)}|}{S_{\mathrm{gt}}^{(n)}} \tag{4.5}$$

ここで N は評価サンプル数である．MAPE は真値が小さい場合に相対誤差が増幅されやすい点に注意が必要であるため，MAE と併用して解釈する．

表 4.1: ポリープ大きさ推定における比較条件

条件	推定手法	NN 補正
条件 1	従来手法	なし
条件 2	従来手法	あり
条件 3	提案手法	なし
条件 4	提案手法	あり

4.2.1　NN 補正

NN 補正は，推定結果に含まれる系統誤差を低減するための回帰モデルである．学習データには愛知医科大学より提供されたフード画像 11 例を用いた．評価の公平性を担保するため，Leave-One-Out Cross Validation（LOOCV）により，各サンプルを 1 例ずつ評価用に保持し，残りの 10 例で学習した補正モデルを適用して評価した．具体的な手順を以下に示す．

1. 11 例のうち 1 例をテストデータとして除外し，残り 10 例で補正モデルを学習する．
2. 学習済み補正モデルを除外した 1 例に適用し，補正後の S_{est} を得る．
3. 上記を全 11 例で繰り返し，すべての推定結果を集約して指標（MAE, MAPE）を算出する．

入力特徴量として，補正前のポリープの大きさに加えて，提案手法ではポリープまでの奥行と実画像におけるポリープ直径のピクセル数を用い，従来手法では推定反射係数を用いた．

4.2.2　誤差比較

表 4.2 に 4 条件の誤差比較を示す．提案手法（条件 3）は従来手法（条件 1）と比較して誤差が低減し，さらに NN 補正を適用した条件 4 により MAE・MAPE がともに改善した．具体的には，提案手法に対する NN 補正の効果として，LOOCV における MAE は 1.561 mm から 0.749 mm へ低下し，MAPE も 39.70%から 17.36%へ低下した．また，従来手法に対しても同様に NN 補正を適用したところ，MAE は 3.789 mm から 1.162 mm へ，MAPE は 73.49%から 23.37%へ低下した．

表 4.2: ポリープ大きさ推定における 4 条件の誤差比較（LOOCV）

条件	MAE [mm]	MAPE [%]
条件 1	3.789	73.49
条件 2	1.162	23.37
条件 3	1.561	39.70
条件 4	0.749	17.36

4.2.3　考察

条件 1 から条件 3 への改善は，従来手法で顕在化していた推定の不安定さに対し，C3VD の真値深度を教師とする回帰により奥行き推定の一貫性が向上したことが寄与したと考えられる．一方で，条件 3 が条件 2 より誤差が大きい点は，C3VD と実画像の撮影環境の差が影響した可能性がある．本研究では焦点距離比によるスケールを導入した結果，内部パラメータ差の一部を低減したものの，照明条件や粘膜反射等に起因する誤差が残存し得る．これに対して条件 4 では，NN 補正が残存する誤差を学習することで最終的なサイズ推定誤差が 0.749 mm まで低減された．

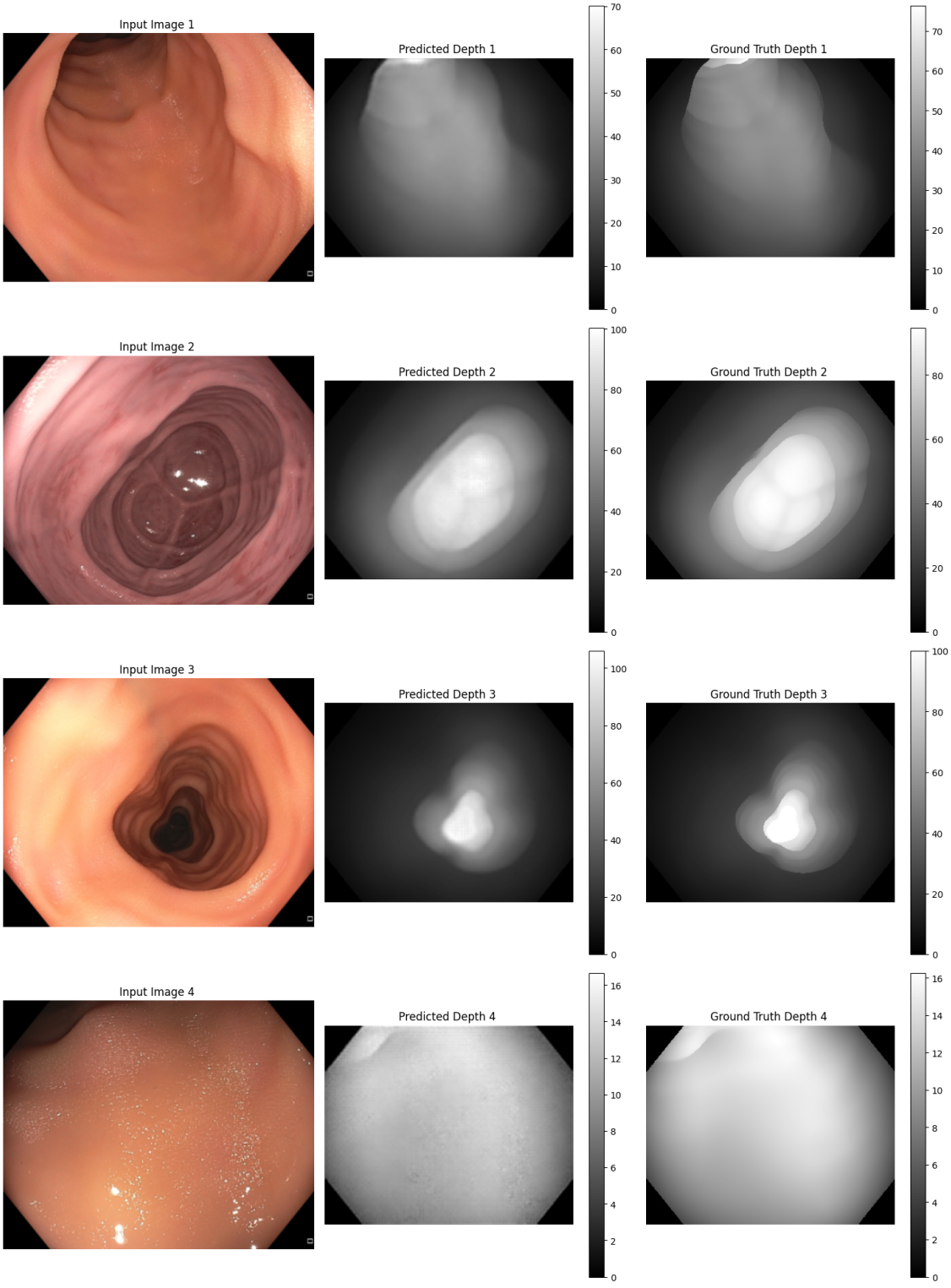


図 4.1: C3VD における深度推定例

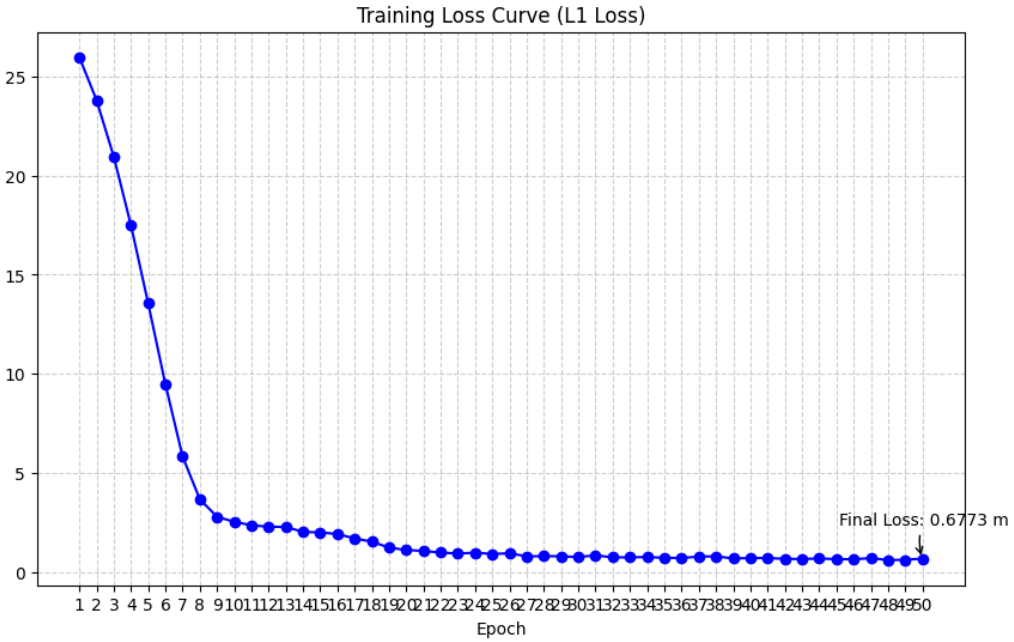
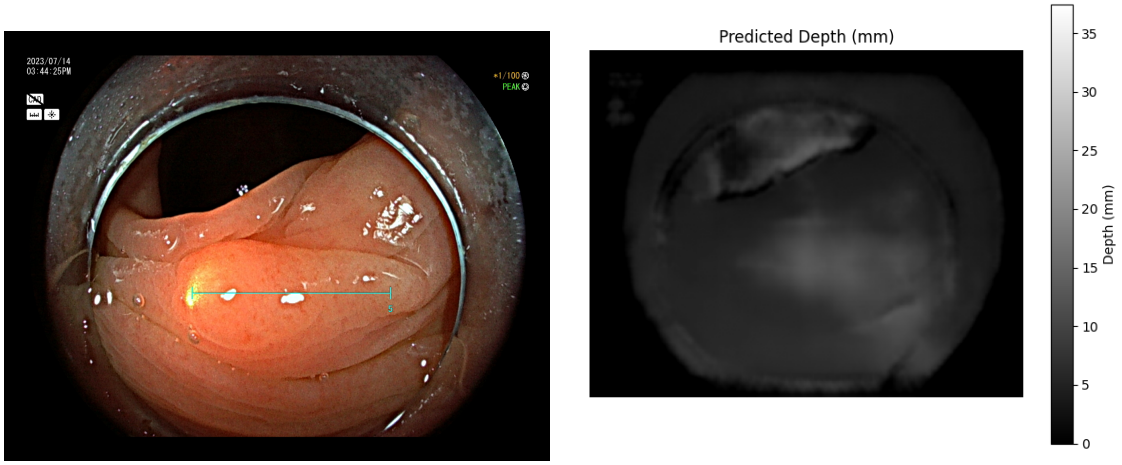
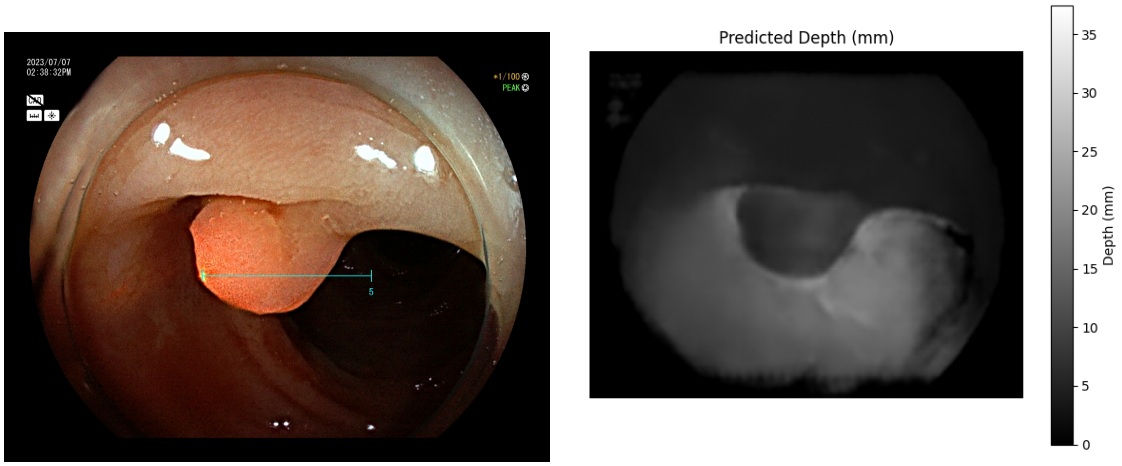


図 4.2: 学習時の損失推移



(a) 実画像（約 3 mm）　(b) 推定深度



(c) 実画像（約 5 mm）　(d) 推定深度

図 4.3: 実画像に対する深度推定結果

第5章　むすび

本研究では，通常の大腸内視鏡画像からポリープの大きさを mm 単位で推定することを目的として，C3VD を教師データとする単眼深度推定と，内視鏡フードの既知寸法を用いたスケール導入，および推定誤差を学習的に補正するニューラルネットワークを組み合わせた手法を提案した．深度推定については，C3VD において真値深度の大域的な分布を概ね再現できることを確認した．また，学習時と異なる光学系で撮影された実画像に対しては，焦点距離比により内視鏡環境下に適したスケールを導入する枠組みを提案し，推定深度を実画像の条件に整合させた上で後段のサイズ算出に用いる手順を示した．ポリープ大きさ推定では，推定深度と画像上のポリープ直径に基づいてサイズを算出し，従来手法との比較により誤差が低減することを確認した．さらに，説明変数に基づく回帰として補正問題を定式化し，ニューラルネットワークによる補正を適用することで最終的な推定精度を改善できることを示した．

今後の課題として，臨床データの拡充と多様な撮影条件下での検証に加え，C3VD と実画像間の差を低減する学習手法の導入が挙げられる．さらに，カメラモデルに基づくスケール推定の高精度化，複数フレーム情報を利用した深度推定の頑健化を進めることで，信頼性の向上が挙げられる．

謝辞

本研究を行うにあたり，終始熱心にご指導をいただいた名古屋工業大学 舟橋健司 准教授，伊藤宏隆 助教に深く感謝いたします．また，日頃よりご指導いただきました中部大学 岩堀祐之 教授（2024 年 12 月 24 日逝去）に深く感謝いたします．心からご冥福をお祈りいたします．

本研究において，内視鏡画像の提供をはじめ，ご協力をいただきました愛知医科大学 春日井邦夫 教授，小笠原尚高 教授に感謝いたします．さらに，本研究を進めるにあたり多くの助言を頂き，御協力いただいた名古屋工業大学 舟橋研究室の皆様に厚く御礼申し上げます．

参考文献

[1] 厚生労働省，“令和 6 年（2024）人口動態統計（確定数）の概況”，https://www.mhlw.go.jp/toukei/saikin/hw/jinkou/kakutei24/index.html（2025/12/25 閲覧）

[2] 大腸癌治療ガイドライン医師用 2019 年版，大腸癌研究会 (編)，金原出版，東京，2019.

[3] K.Hasegawa, Y.Sato, “Endoscope system for high-speed 3D measurement”, Systems and Computers in Japan, pp.30-39, 2001.

[4] Y.S. Hyun, D.S. Han, J.H.Bae, H.S.Park, C.S. Eun, “Graduated injection needles and snares for polypectomy are useful for measuring colorectal polyp size”, Digestive and Liver Disease, vol.43, no.5, pp.391-394, 2011.

[5] A.M. Kaz, A. Anwar, D.R. O’Neill, J.A. Dominitz, “Use of a novel polyp ‘ruler snare’ improves estimation of colon polyp size”, Gastrointestinal Endoscopy, vol.83, no.4, pp.812–816, 2016.

[6] B.K.P. Horn, “Obtaining shape from shading information”, MIT Press, pp.123-171, 1989.

[7] P.W. Verbeek, B.J.H. Verwer, “Shading from shape, the eikonal equation solved by gray-weighted distance transform”, Pattern Recognition Letters, vol.63, pp.681–690, 1990.

[8] J.A. Sethian, “A Fast Marching Level Set Method for Monotonically Advancing Fronts”, Proceedings of the National Academy of Sciences of the United States of America, vol.93, no.4, pp.1591–1593, 1996.

[9] R. Kimmel, J.A. Sethian, “Optimal Algorithm for Shape from Shading and Path Planning”, Journal of Mathematical Imaging and Vision, vol.14, no.3, pp.237–244, 2001.

[10] A. Tankus, N. Sochen, Y. Yeshurun, “Shape-from-Shading Under Perspective Projection”, International Journal of Computer Vision, vol.63, no.1, pp.21-43, 2005.

[11] S.Y. Yuen, Y.Y. Tsui, C.K. Chow, “A Fast Marching Formulation of Perspective Shape from Shading under Frontal Illumination”, Pattern Recognition Letters, vol.28, no.7, pp.806-824, 2007.

[12] Y. Iwahori, K. Iwai, R.J. Woodham, H. Kawanaka, S. Fukui, K. Kasugai, “Extending Fast Marching Method under Point Light Source Illumination and Perspective Projection”, ICPR2010, pp.1650-1653, 2010.

[13] Y Iwahori, K Tatematsu, T Nakamura, S Fukui , RJ Woodham, K Kasugai, “3D Shape Recovery from Endoscope Image Based on Both Photometric and Geometric Constraints”, Knowledge-Based Information Systems in Practice. Smart Innovation, Systems and Technologies, vol.30, pp.65-80, 2015.

[14] H. Usami, Y. Iwahori, M.K. Bhuyan, A. Wang, N. Ogasawara, K. Kasugai, “Recovering Polyp Shape from an Endoscope Image Using Two Light Sources”, International Journal of Software Innovation, vol.5, no.2, pp.33-54, 2017.

[15] Y. Iwahori, S. Miyazaki, H. Usami, Y. Hanai, B. Kijsirikul, K. Kasugai, “Automatic Detection of LST-Type Polyp by CNN Using Depth Map”, Handbook of Artificial Intelligence in Healthcare, pp.177-196, 2021.

[16] Y. Iwahori, T. Suda, K. Funahashi, H. Usami, A. Wang, M.K. Bhuyan, K. Kasugai, “Shape Recovery of Polyp from Endoscope Image Using Blood Vessel Information”, Computational Science/Intelligence and Applied Informatics, pp.165-184, 2017.

[17] H. Usami, Y. Iwahori, M.K. Bhuyan, A. Wang, N. Ogasawara, K. Kasugai, “Polyp Shape Recovery using Vascular Border from Single Colonoscopy Image”, BIOIMAGING 2019, pp.104-111, 2019.

[18] 江本　峻，“単一内視鏡画像からのポリープ形状復元と血管情報を利用したその大きさ推定”，名古屋工業大学大学院修士論文，2022.

[19] R. Kikuchi, Y. Iwahori, K. Funahashi, M.K. Bhuyan, A. Wang, N. Ogasawara, K. Kasugai, “Polyp Size and Shape Estimation by Using an Endoscopic Hood Information”, Procedia Computer Science, vol.225, pp.2412-2419, 2023.

[20] N. Ikeda, H. Usami, Y. Iwahori, B. Kijsirikul, K. Kasugai, “Generating Lambertian Image by Removing Specular Reflection Component and Difference of Reflectance Factor Using HSV”, ITC-CSCC 2016, pp.547-550, 2016.

[21] “Colonoscopy 3D Video Dataset (C3VD)”, https://durrlab.github.io/C3VD/ (2026/1/10 閲覧)

[22] T.L. Bobrow, M. Golhar, R. Vijayan, V.S. Akshintala, J.R. Garcia, N.J. Durr, “Colonoscopy 3D video dataset with paired depth from 2D-3D registration”, Medical Image Analysis, vol.90, p.102956, 2023.

発表論文リスト

1. 菊地 遼，岩堀 祐之，舟橋 健司，小笠原 尚高，春日井 邦夫，“内視鏡フードの利用によるポリープの 3 次元情報と大きさの復元”，第 20 回情報学ワークショップ（WiNF）抄録集，S-4B-7，2022.
2. 菊地 遼，岩堀 祐之，舟橋 健司，小笠原 尚高，春日井 邦夫，“内視鏡フード画像からのポリープの形状及び大きさの推定”，第 26 回画像の認識・理解シンポジウム (MIRU) 予稿集，IS1-94，2023.
3. R. Kikuchi, Y. Iwahori, K. Funahashi, M.K. Bhuyan, A. Wang, N. Ogasawara, K. Kasugai, “Polyp Size and Shape Estimation by Using an Endoscopic Hood Information”, Procedia Computer Science, vol.225, pp.2412-2419, 2023.
4. 菊地 遼，岩堀 祐之，舟橋 健司，小笠原 尚高，春日井 邦夫，“内視鏡フードを用いたポリープの大きさ推定及び NN による誤差補正”，第 28 回 画像の認識・理解シンポジウム（MIRU）予稿集，IS3-020，2025.