Метод построения скелета изображения

Ван Сыжэнь

Аспирант

МЕХМАТ МГУ, Москва, Россия

E-mail: sirenexcelsior@gmail.com

***Научный руководитель** — Профессор В.Н.Козлов*

Обнаружение локальных признаков является ключевой задачей визуальной локализации и трёхмерной реконструкции, однако существующие методы глубокого обучения не обеспечивают математических гарантий в условиях экстремальных аффинных преобразований. В настоящей работе предлагается гибридная архитектура, внедряющая аффинно-инвариантное кодирование из работ В.Н.Козлова для обеспечения доказательных геометрических ограничений. Метод основан на вычислении инвариантов отношений площадей треугольников для скелета ключевых точек и регуляризации согласованности посредством дифференцируемых гистограмм. Предложенный подход сохраняет дискриминативную способность дескрипторов и одновременно существенно повышает устойчивость модели к сложным аффинным преобразованиям.

В настоящей работе модель R2D2 используется в качестве основы для формирования скелета ключевых точек, посредством полностью сверточной сети одновременно предсказывая их повторяемость и надежность. С точки зрения дискретной геометрии, извлекаемое множество ключевых точек $P = \{p_1, \dots, p_K\}$ представляет собой аппроксимацию конечного множества точек в евклидовом пространстве. В. Н. Козлов доказал, что код изображения [1], основанный на отношениях площадей треугольников, обладает строгой инвариантностью при аффинных преобразованиях. Геометрический инвариант определяется следующим образом:

$$\rho_{ijk,lmn} = \frac{S_{ijk}}{S_{lmn}}, \quad S_{lmn} \neq 0$$

где S_{ijk} и S_{lmn} — площади треугольников, образованных произвольными тройками точек; иначе говоря, отношение площадей любых треугольников сохраняется при аффинных преобразованиях.

Дифференцируемая реализация. Для построения функции потерь, пригодной для оптимизации модели глубокого обучения, в настоящей работе используется концепция «стертого кода» [3], сохраняющая множество инвариантов $\mathcal{Q}(P) = \{\rho_1, \dots, \rho_N\}$, где ρ_s обозначает s-й вычисленный инвариант, а процесс укрупнения аппроксимируется посредством дифференцируемых гистограмм. b-я компонента вектора укрупненного кода $V \in \mathbb{R}^B$ (где B — число бинов гистограммы) вычисляется следующим образом:

$$V_b = \sum_{s=1}^{N} \exp\left(-\frac{(\rho^{(s)} - c_b)^2}{2\sigma^2}\right),$$

где N — количество выбранных инвариантов в выборке, $\rho^{(s)}$ — значение s-го инварианта, c_b — центр b-го бина, а σ контролирует степень сглаживания. Данная дифференцируемая конструкция поддерживает обратное распространение ошибки, позволяя напрямую интегрировать геометрические ограничения в процесс глубокого обучения.

Согласно неравенствам концентрации в теории статистического обучения, для гарантии того, что ошибка между эмпирическим и истинным распределениями будет меньше ϵ (при доверительной вероятности $1-\delta$), количество ключевых точек K должно удовлетворять теоретической нижней границе:

$$K \geq \sqrt[3]{\frac{3}{\alpha\epsilon^2} \ln\left(\frac{2B}{\delta}\right)}$$

где α — частота выборки. Данная граница нечувствительна к изменениям доверительной вероятности: повышение доверительной вероятности с 95% до 99% требует увеличения количества ключевых точек K лишь примерно на 20%. Следовательно, на практике выбор среднего уровня доверительной вероятности достаточен для обеспечения статистической устойчивости при сохранении вычислительной эффективности.

Аффинно-инвариантная регуляризация. Для обеспечения геометрических ограничений аффинной инвариантности вводится аффинно-инвариантная регуляризационная функция потерь, оценивающая согласованность геометрической структуры между различными видами. Она объединяется с исходными функциями потерь R2D2, формируя гибридную целевую функцию оптимизации. Аффинно-инвариантная регуляризационная функция потерь определяется как расстояние между векторами укрупненного кода для пары аффинно-эквивалентных изображений:

$$L_{Kozlov}(I, I') = \|V - V'\|_2^2$$

Итоговая общая функция потерь определяется следующим образом:

$$L_{total} = L_{rep} + L_{AP\kappa} + \lambda \cdot L_{Kozlov}$$

где L_{rep} и $L_{AP\kappa}$ обозначают исходные функции потерь метода R2D2[2], а λ — коэффициент балансировки.

Затем модель оптимизируется в режиме сквозного обучения, накладывая жёсткие геометрические ограничения на пространство представлений, что гарантирует сохранение структурной согласованности скелета ключевых точек при произвольных аффинных преобразованиях. Таким образом достигается баланс между строгой инвариантностью, обусловленной теорией дискретной геометрии, и адаптивностью данных, что повышает надёжность сопоставления в условиях экстремальных геометрических искажений.

Литература

1. Kozlov V. N. Recognition of Images Represented by a Final Set of Points // Journal of Mathematical Sciences. 2011. Vol. 172. No. 5. P. 690-699.
2. Revaud J., De Souza C., Humenberger M., Weinzaepfel P. R2d2: Reliable and repeatable detector and descriptor // Advances in Neural Information Processing Systems. 2019. Vol. 32.
3. Kozlov V. N. An Analog of the Electronic Digital Signature Based on Codes Defining Images with Accuracy up to Affine Transforms // Pattern Recognition and Image Analysis. 2023. Vol. 33. No. 2. P. 203-207.