Rethinking Generalization Performance of Surgical Phase Recognition with Expert-Generated Annotations

Seungbum Hong, Jiwon Lee, Bokyung Park, and Min-Kook Choi[*]
hutom
Seoul, Republic of Korea
`{qbration21, jiwon, bokyung, mkchoi}@hutom.io`

Woo Jin Hyung
hutom, Yonsei University College of Medicine
Seoul, Republic of Korea
`wjhyung@hutom.io`

Ahmed A. Alwusaibie, Anwar H. Alfadhel, and SungHyun Park
Yonsei University College of Medicine
Seoul, Republic of Korea
`GODON@yush.ac`

Abstract

As the area of application of deep neural networks expands to areas requiring expertise, e.g., in medicine and law, more exquisite annotation processes for expert knowledge training are required. In particular, it is difficult to guarantee generalization performance in the clinical field in the case of expert knowledge training where opinions may differ even among experts on annotations. To raise the issue of the annotation generation process for expertise training of CNNs, we verified the annotations for surgical phase recognition of laparoscopic cholecystectomy and subtotal gastrectomy for gastric cancer. We produce calibrated annotations for the seven phases of cholecystectomy by analyzing the discrepancies of previously annotated labels and by discussing the criteria of surgical phases. For gastrectomy for gastric cancer has more complex twenty-one surgical phases, we generate consensus annotation by the revision process with five specialists. By training the CNN-based surgical phase recognition networks with revised annotations, we achieved improved generalization performance over models trained with original annotation under the same cross-validation settings. We showed that the expertise data annotation pipeline for deep neural networks should be more rigorous based on the type of problem to apply clinical field.

1 Introduction

Because algorithms using pattern recognition or machine learning are applied for various tasks, problem definition and data acquisition are as important as the model's performance [1-4]. Training using deep neural networks, a trend in machine learning, requires large amounts of annotated data to avoid overfitting with a large number of parameters. Crowdsourcing-based tools, such as Amazon

---

[*]Corresponding author.