ترجمه کامل:

(ELSEVIER)

نشریه پردازش و کنترل سیگنال‌های زیست‌پزشکی ۸۸ (۲۰۲۴) ۱۰۵۶۳۷

مبتنی بر نمونه‌کاهی انطباقی و کلیپ‌های کوتاه CNN-LSTM مدل دو جریانی برای تشخیص فاز جراحی در ویدئوهای کوله‌سیستکتومی (برداشتن کیسه صفرا)

نویسندگان: سانگ-گو لی، گا-یانگ کیم، یو-نا هوانگ، جی-یون کوون، سونگ-مین کیم

دپارتمان تجهیزات پزشکی و سلامت، دانشگاه دونگ‌گوک-سئول، جمهوری کره

چکیده:

ناشی از (Overfitting) تشخیص فازهای جراحی به دلیل وجود مشکل بیش‌برازش میان فازهای مختلف جراحی، امری (Imbalanced Data) عدم تعادل داده‌ها مبتنی بر نرخ (Undersampling) چالش‌برانگیز است. ما یک روش نمونه‌کاهی نمونه‌برداری انطباقی پیشنهاد کرده‌ایم که می‌تواند تعداد داده‌های هر فاز جراحی را به صورت مشابه تولید کند تا از یادگیری سوگیرانه جلوگیری شود. برای بهبود نیز معرفی کردیم که CNN-LSTM عملکرد روش خود، یک مدل دو جریانی .می‌تواند اطلاعات زمانی تغییرات رفتاری بین هر فریم تصویر را استخراج کند

ابتدا، در مجموع ۴۰,۲۳۶ کلیپ کوتاه با استفاده از نرخ زیرنمونه‌برداری انطباقی از کل ویدئو استخراج شد. هر کلیپ کوتاه وارد یک شبکه از پیش آموزش‌دیده GoogLeNet شد. خروجی حاوی اطلاعات بصری بلافاصله به یک مدل LSTM تغذیه شد تا اطلاعات زمانی فریم‌های (Sequence-to-Sequence) توالی‌به-توالی توالی‌به-بردار LSTM همسایه را در یک کلیپ کوتاه استخراج کند. هم‌زمان، از یک دیگر برای استخراج اطلاعات زمانی از تمام فریم‌های (Sequence-to-Vector) .تصویری متوالی جهت پیش‌بینی فاز نهایی جراحی استفاده شد

ارزیابی شد و با کسب امتیاز Cholec80 روش پیشنهادی با مجموعه داده عمومی معادل ۹۸.۰۰٪ از روش‌های پیشین پیشی AUC معادل ۸۷.۱۲٪ و شاخص F1 بین تمامی فازها پس از اعمال نمونه‌کاهی F1 گرفت. علاوه بر این، انحراف امتیاز حدود ۱۰٪ کاهش یافت. نتایج تجربی تایید کرد که استفاده از روش پیشنهادی ما

می‌تواند اطلاعات زمانی غنی را از کلیپ‌های کوتاه یاد بگیرد و از معماری سنتی CNN-LSTM تک‌جریانی عملکرد بهتری ارائه دهد.

کلمات کلیدی: تشخیص خودکار فاز جراحی، کوله‌سیستکتومی، ویدئوی آندوسکوپی، نمونه‌کاهی، CNN-LSTM مبتنی بر کلیپ کوتاه، مدل دو جریانی.

مقدمه ۱.

روشی است که با کاهش ناحیه برش در حین لاپاراتومی، (MIS) جراحی کم‌تهاجمی به دلیل مزایای MIS. زخم‌ها یا اسکارها را بر روی بدن انسان به حداقل می‌رساند متعدد مانند خونریزی و اسکار کمتر، دوره بهبودی کوتاه و خطر عفونت کمتر در حال رایج شدن است. یک جراحی کم‌تهاجمی نمونه، کوله‌سیستکتومی (عمل برداشتن کیسه صفرا) است. حدود ۹۰ درصد کوله‌سیستکتومی‌ها در ایالات متحده به روش لاپاروسکوپی انجام می‌شوند. با افزایش علاقه به جراحی‌های کم‌تهاجمی، تمایل به تشخیص جریان کار جراحی از روی داده‌های ویدئویی لاپاروسکوپی نیز رو به افزایش است.

تشخیص جریان کار جراحی، به ویژه برای تشخیص فاز، تشخیص ابزار و تشخیص رویدادها، یک عامل کلیدی برای حمایت از تیم‌های جراحی و بهینه‌سازی درمان بیمار در داخل اتاق عمل است. در عمل کوله‌سیستکتومی، تشخیص فاز جراحی برای بهینه‌سازی فرآیند جراحی، آموزش جراحان کم‌تجربه و ارائه کمک‌های حین عمل بسیار مهم است. تشخیص فاز جراحی می‌تواند مراحل مختلف یک فرآیند کامل جراحی کوله‌سیستکتومی را شناسایی کند؛ مراحلی شامل: آماده‌سازی (Preparation)، دایسکشن مثلث کالوت (Calot's triangle dissection)، برش (Gallbladder dissection)، دایسکشن کیسه صفرا (Clipping cutting)، و کلمپ کردن (Gallbladder packaging)، بسته‌بندی کیسه صفرا (Cleaning coagulation)، و خارج کردن کیسه صفرا (retraction) انعقاد.

کارهای مرتبط ۲.

و شبکه عصبی مصنوعی رایج‌ترین مدل‌های یادگیری (HMM) مدل پنهان مارکوف ماشین مورد استفاده هستند. مطالعات قبلی برای تشخیص فاز جراحی، استفاده از منبع سیگنال مبتنی بر حسگر را به عنوان داده ورودی بررسی کرده‌اند. اکثر مطالعات قبلی از داده‌های ویدئویی یا سیگنال‌های باینری ابزارها استفاده کرده‌اند. دو مدل رایج برای تشخیص فاز از HMM و (DTW) مدل‌های تبرب زمانی پویا سلسله‌مراتبی با سایر HMM روی سیگنال‌های باینری ابزار جراحی بودند. مدل

ترکیب شده (CNN) مدل‌های یادگیری ماشین مانند شبکه‌های عصبی کانولوشنی است تا دقت را بهبود بخشد. با این حال، این سیگنال‌ها معمولاً از طریق نشانه‌گذاری دستی ویدئوهای جراحی یا با نصب حسگرهای اضافی به دست می‌آیند.

بسیاری از مطالعات بر استفاده از ویژگی‌های بصری با ادغام شبکه‌های عصبی یا حافظه طولانی (RNN) با شبکه‌های عصبی بازگشتی (CNN) کانولوشنی تمرکز کرده‌اند. اگرچه در مطالعات قبلی از ویژگی‌های بصری (LSTM) کوتاه‌مدت همراه با اطلاعات ابزار جراحی استفاده شده است، اما برچسب‌گذاری دستی زمان‌بر از RNN است که یک عیب محسوب می‌شود. برای کاهش زمان آموزش، به جای و شبکه ترنسفورمر استفاده شده است. با این حال، (TCN) شبکه کانولوشنی زمانی کارهای موجود در زمینه تشخیص فاز جراحی از ویدئوهای حین عمل به دلیل مشکلاتی مانند تفاوت در آناتومی بیماران، سبک‌های مختلف جراحان و عدم تعادل کلاس‌ها همچنان چالش‌برانگیز است که می‌تواند منجر به کاهش عملکرد شود. تا استفاده Cholec80 جایی که ما می‌دانیم، هیچ‌یک از مطالعاتی که از مجموعه داده کرده‌اند، به مشکل عدم تعادل کلاس نپرداخته‌اند.

هدف این مطالعه تشخیص فاز در ویدئوهای لاپاروسکوپی برای مراحل جراحی مبتنی بر زیرنمونه‌برداری CNN-LSTM کوله‌سیستکتومی بود. ما مدل دو جریانی زمانی انطباقی را با یادگیری روابط زمانی بین توالی‌های طولانی و کلیپ‌های کوتاه معرفی کردیم. نوآوری‌های اصلی ما دو مورد است: (۱) حل مشکل عدم تعادل که از نرخ‌های (Undersampling) کلاس‌ها با یک روش جدید نمونه‌کاهی برای هر فاز استفاده می‌کند؛ و (۲) (fps فریم بر ثانیه یا) نمونه‌برداری متفاوت برای GoogLeNet که در آن از CNN-LSTM معرفی معماری دو جریانی برای استخراج اطلاعات LSTM استخراج ویژگی‌های بصری و از دو جریان زمانی استفاده می‌شود.

مواد و روش‌ها ۳.

یک نمای کلی از روش‌های پیشنهادی و فرآیند آموزش ما در شکل ۱ ارائه شده است. ما یک استراتژی نمونه‌کاهی انطباقی پیشنهاد کردیم که می‌تواند یادگیری سوگیرانه را با ایجاد تعداد مشابهی کلیپ کوتاه بین هر ۷ کلاس کاهش دهد. ما همچنین از یک برای استخراج ویژگی‌های بصری بهره بردیم و از دو GoogLeNet شبکه برای بهینه‌سازی اطلاعات زمانی استفاده کردیم LSTM.

مجموعه داده ۳.۱.

شامل ۸۰ ویدئوی کوله‌سیستکتومی Cholec80 این مطالعه از مجموعه داده عمومی استفاده کرد که توسط ۱۳ جراح انجام شده بود. تمامی ویدئوها با نرخ ۲۵ فریم بر ثانیه و رزولوشن ۱۰۸۰×۱۹۲۰ یا ۴۸۰×۸۵۴ پیکسل ضبط شده بودند. جدول ۱ نمای کلی داده‌ها را با تعداد کل فریم‌ها و مدت زمان کل برای هر فاز نشان می‌دهد.

جدول ۱: ترکیبات مجموعه داده Cholec80

- فاز ۱ (آماده‌سازی): ۲۱۴,۳۰۰ فریم (۸,۵۷۲ ثانیه)
- فاز ۲ (دایسکشن مثلث کالوت): ۱,۸۴۲,۵۲۵ فریم (۷۳,۷۰۱ ثانیه)
- فاز ۳ (برش و کلمپ): ۳۵۲,۰۰۰ فریم (۱۴,۰۸۰ ثانیه)
- فاز ۴ (دایسکشن کیسه صفرا): ۱,۴۶۰,۸۲۵ فریم (۵۸,۴۳۳ ثانیه)
- فاز ۵ (بسته‌بندی کیسه صفرا): ۱۹۰,۴۵۰ فریم (۷,۶۱۸ ثانیه)
- فاز ۶ (پاک‌سازی و انعقاد): ۴۸۵,۶۵۰ فریم (۱۹,۴۲۶ ثانیه)
- فاز ۷ (خارج کردن کیسه صفرا): ۱۶۵,۹۲۵ فریم (۶,۶۳۷ ثانیه)
- **مجموع:** ۴,۷۱۱,۶۷۵ فریم (۱۸۸,۴۶۷ ثانیه)

۳.۲. روش نمونه‌کاهی مبتنی بر نرخ نمونه‌برداری انطباقی

فازهای ۲ و ۴ حدود ۶۸ درصد از کل کلیپ‌های مجموعه آموزش را تشکیل می‌دادند. این عدم تعادل داده‌ها می‌تواند منجر به یادگیری سوگیرانه شود. برای حل این مشکل، ما کل ویدئو را به کلیپ‌های کوتاهی تقسیم کردیم که شامل ۱۰ فریم با نمونه‌برداری (Stride) نرخ‌های نمونه‌برداری متفاوت برای هر فاز بود. گام حرکت برای هر فاز با استفاده از معادله (۱) محاسبه شد:

$$stride = \lfloor MP \times 25 \rfloor$$

بالاترین تعداد فریم در بین تمامی M تعداد فریم‌های تصویر در هر فاز و P که در آن فازها (در اینجا فاز ۲) است. به این ترتیب برای فازهای ۱ تا ۷ به ترتیب نرخ‌های ۸.۳۳، ۱، ۵، ۱.۲۵، ۸.۳۳، ۴.۱۷ و ۱۲.۵ فریم بر ثانیه به دست آمد تا در نهایت ۴۰,۲۳۶ کلیپ کوتاه و متعادل تولید شود.

۳.۳. معماری دو جریانی CNN-LSTM

ما جایگزین LSTM با مدل دو جریانی GoogLeNet آخرین لایه طبقه‌بندی‌کننده Global توسط یک لایه GoogLeNet شد. خروجی‌های آخرین ماژول اینسپشن یک مدل توالی‌به- LSTM برداری شدند. یکی از جریان‌های Max Pooling 2D بود که اطلاعات بلندمدت را دریافت می‌کرد و جریان دیگر یک (seq2vec) بردار بود که اطلاعات کوتاه‌مدت درون هر کلیپ را (seq2seq) مدل توالی‌به-توالی استخراج می‌کرد. در نهایت، امتیازهای پیش‌بینی این دو جریان با یک تابع جمع ترکیب شدند (Add).

:هایپرپارامترهای شبکه

- بهینه‌ساز: Adam (۰.۹ :مومنتم)
- معادل ۰.۲۵ :LSTM نرخ دراپ‌اوت لایه
- نرخ یادگیری اولیه: ۰.۰۰۰۰۱ (که در اپوک ۱۵ نصف می‌شود)
- تعداد اپوک‌ها: ۴۵
- تابع فعال‌سازی LSTM: سیگموئید (Sigmoid)
- تابع زیان: Categorical Cross-Entropy
- اندازه بچ (Batch Size): ۱۶

۴. نتایج تجربی

- **۴.۱. مقایسه مدل‌های CNN:** مدل GoogLeNet با کسب امتیاز F1 بالاتر بهترین (۷۶.۰۲٪) ResNet50 و (۷۵.۲۴٪) VGG16 (۷۶.۷۹٪) نسبت به عملکرد را داشت.
- **۴.۲. تعداد فریم‌های کلیپ کوتاه:** تعداد ۱۰ فریم در هر کلیپ بالاترین امتیاز را ثبت کرد (۹۶.۳۳٪) AUC و بالاترین میزان (۷۶.۷۹٪) F1.
- **۴.۳. تاثیر نمونه‌کاهی انطباقی:** پس از اعمال نمونه‌کاهی انطباقی (گروه ۳)، به ۹۸.۰۰٪ افزایش AUC کل به طرز چشمگیری به ۸۷.۱۲٪ و F1 امتیاز یافت و انحراف میان فازها ۱۰٪ کاهش پیدا کرد.
- **۴.۴. مقایسه با کارهای قبلی:** مدل پیشنهادی با امتیاز F1 ۸۷.۱۲٪ از پیشی (۸۳.۴۰٪) TeCNO و (۸۴.۴۹٪) OperA مدل‌های مطرحی مانند گرفت (جدول ۶).

بحث و ۶. نتیجه‌گیری ۵.

را همراه با روش نمونه‌کاهی مبتنی بر CNN-LSTM ما یک مدل دو جریانی کلیپ‌های کوتاه برای تشخیص فاز جراحی در ویدئوهای کوله‌سیستکتومی پیشنهاد کردیم. روش پیشنهادی ما توانست یادگیری سوگیرانه را با متعادل‌سازی داده‌ها به سبب شد تا اطلاعات مربوط به LSTM حداقل برساند. ساختار موازی دو جریان تغییرات حرکتی بین فریم‌ها و همچنین کل فعالیت درون کلیپ کوتاه به خوبی حفظ و یاد گرفته شود و مدل بتواند فازهای جراحی را به صورت زنده و نزدیک به زمان واقعی (Near Real-time) پیش‌بینی کند.