DOI: 10.48305/jims.v43.i836.1383

مجله دانشکده پزشکی اصفهان — سال چهل و سوم/شماره‌ی ۸۳۶/هفته‌ی چهارم آذر ۱۴۰۴

تاریخ دریافت: ۱۴۰۳/۰۹/۱۹ — تاریخ پذیرش: ۱۴۰۴/۰۹/۰۱ — تاریخ چاپ: ۱۴۰۴/۰۹/۲۹

تخمین ضربان قلب به صورت غیرتماسی مبتنی بر یادگیری عمیق

سارا پورمحمدی[۱]، کیان شاهی[۲]

مقاله پژوهشی

چکیده

مقدمه: با استفاده از تکنولوژی فوتوپلتیسموگرافی غیرتماسی (remote Photoplethysmography) rPPG پایش ضربان قلب (Heart rate) HR بدون نیاز به تماس فیزیکی امکان‌پذیر شده است. در دهه‌های گذشته، روش‌هایی برای تخمین سیگنال rPPG و نرخ ضربان قلب با استفاده از فریم‌های ویدئویی توسعه داده شده‌اند. در سال‌های اخیر روش‌های یادگیری عمیق نیز در این حوزه به کار گرفته شده و عملکرد مناسبی داشته‌اند.

روش‌ها: در این مطالعه، از یک شبکه عصبی عمیق (MTTS-CAN) همراه با الگوریتم تشخیص چهره (mediapipe) به منظور تخمین ضربان قلب به صورت غیر تماسی از روی ویدئوی چهره افراد در یک مجموعه داده عمومی (PURE) و یک مجموعه داده بومی (Stroop) بهره گرفته شده است.

یافته‌ها: نتایج پیاده‌سازی بر روی مجموعه داده PURE قابل مقایسه با نتایج منتشر شده در مقالات (MAE:7.72bpm) است. بعلاوه نتایج بر روی مجموعه داده بومی نیز قابل قبول (MAE:5.53bpm) می‌باشد.

نتیجه‌گیری: در این مطالعه، دقت تخمین قابل قبولی برای ضربان قلب غیرتماسی ارائه شده است. علاوه بر آن، نتایج نشان می‌دهند که روش پیشنهادی به دیتاست خاصی وابسته نبوده و در یک دیتاست بومی در مقایسه با دیتاست بنچمارک نتایج مناسبی ایجاد کرده است.

واژگان کلیدی: ضربان قلب؛ پایش ضربان قلب غیرتماسی؛ سیگنال rPPG؛ دیتاست PURE؛ یادگیری عمیق

ارجاع: پورمحمدی سارا، شاهی کیان. **تخمین ضربان قلب به صورت غیرتماسی مبتنی بر یادگیری عمیق.** مجله دانشکده پزشکی اصفهان ۱۴۰۴؛ ۴۳ (۸۳۶): ۱۳۸۳– ۱۳۸۹.

مقدمه

علائم حیاتی شامل ضربان قلب، نرخ تنفس، فشارخون، دمای بدن و سطح اکسیژن خون اطلاعاتی از سیستم قلبی– عروقی و تنفسی فرد ارائه می‌دهند. اندازه‌گیری این علائم حیاتی به منظور پایش وضعیت سلامتی و دنبال کردن مشکلات جسمی یا روحی فرد ضروری است. با پیشرفت تکنولوژی سنسورهای پوشیدنی، امروزه پایش علائم حیاتی خارج از بیمارستان‌ها و مراکز درمانی کار ساده‌ای است. اما پایش غیرتماسی علائم حیاتی دریچه‌ی تازه‌ای بر روی خدمات سلامت الکترونیک گشوده است. استفاده از ابزارهای غیر تماسی متنوعی مانند دوربین های حرارتی، رادارهای ماکروویو و دوربین های مرئی در حال توسعه هستند، اما در این میان استفاده از تصاویر دوربین مرئی ارزان، غیر تهاجمی، غیر مزاحم و به طور گسترده در دسترس است.

ایده‌ی استفاده از تصاویر دوربین مرئی بر این اساس است که خون جاری در سیستم قلبی– عروقی همراه با فازهای انقباض و استراحت قلب، در حال تغییر است. این جریان خون ضربان‌دار با درجات مختلف در قسمت‌های مختلف بدن قابل تشخیص است و باعث تغییرات نامحسوسی در رنگ پوست می‌شود که توسط دوربین‌های ثبت تصویر قابل تشخیص است (۱). کاربردهای پایش ضربان قلب بر مبنای تصاویر دوربین کلینیکی مانند پایش نوزادان (۲)، بیماران در ICU (۳)، بیماران در هنگام MRI (٤) و یا قبل، بعد وحین عمل جراحی (٤) است.

کاربردهای غیر کلینیکی آن نیز مانند پایش بیماران و سالمندان در خانه (۵)، پایش راننده و مسافران در وسیله نقلیه و پایش استرس (٦) است. بیشتر تحقیقات قبلی تشخیص ضربان قلب غیر تماسی بر

۱- دکتری، گروه پردازش علائم زیستی، پژوهشکده پردازش سیگنال، پژوهشگاه توسعه فناوری‌های پیشرفته، تهران، ایران

۲– کارشناسی ارشد، گروه پردازش علائم زیستی، پژوهشکده پردازش سیگنال، پژوهشگاه توسعه فناوری‌های پیشرفته، تهران، ایران

نویسنده‌ی مسؤول: سارا پورمحمدی: دکتری، گروه پردازش علائم زیستی، پژوهشکده پردازش سیگنال، پژوهشگاه توسعه فناوری‌های پیشرفته، تهران، ایران

Email: Pourmohammadi@semnan.ac.ir

تخمین سیگنال فوتوپلتیسموگرافی غیرتماسی (rPPG) با استفاده از تصاویر دوربین مرئی متمرکز شده‌اند (۷).

به طور کلی در همه‌ی روش‌های سنتی تخمین ضربان قلب از تصاویر دوربین مرئی، ابتدا تصاویر با یک دوربین دیجیتال ثبت شده (یا رشته تصاویر به صورت آنلاین ارسال می‌شوند)، سپس یک الگوریتم تشخیص چهره مانند Viola-Jones (۸)، برای به دست آوردن مختصات مرزی صورت اعمال می‌شود. پس از این کار، مناطق مورد علاقه (ROI (Region of interest مانند گونه‌ها یا پیشانی انتخاب شده تا ناحیه‌ای حاوی سیگنال قوی به دست آید. پیکسل‌های درون ROI(ها) برای استخراج سیگنال rPPG استفاده می‌شوند و ضربان قلب با پس پردازش بیشتر، که معمولاً شامل تجزیه و تحلیل فرکانس و تشخیص پیک می‌شود، تخمین زده می‌شود.

در این مقاله، ضربان قلب با یک تکنیک ترکیبی با استفاده از الگوریتم تشخیص چهره و روش شبکه‌ی عصبی کانوولوشنال سه بعدی عمیق با نام MTTS-CAN تخمین زده شده است. ورودی شبکه‌ی عمیق، فریم‌های ویدئویی پردازش شده با روش تشخیص چهره و خروجی آن سیگنال rPPG است. پس از آن با استفاده از یک مرحله پس پردازش، ضربان قلب از rPPG، استخراج شده است. اگرچه روش MTTS-CAN در (۹) معرفی شده و از آن برای تخمین ضربان قلب بهره برده‌اند، نوآوری این مقاله پیشنهاد استفاده از یک روش تشخیص چهره به همراه روش یادگیری عمیق برای افزایش کارآیی روش پیشنهادی است. همچنین الگوریتم پیشنهادی بر روی یک دیتاست بومی با کیفیت تصاویر معمولی (شرایط نور معمولی اتاق و آزمودنی‌های بومی در حین انجام یک تسک شناختی) پیاده شده و نتایج آن با یک دیتاست عمومی مقایسه شده است. هدف نهایی نویسندگان، استفاده‌ی کاربردی و برخط از الگوریتم پیشنهادی در یک سامانه پایش ضربان قلب/ برانگیختگی غیر تماسی است.

ادامه این مقاله به ترتیب زیر سامان یافته است: در بخش دوم به معرفی دیتاست‌ها و توصیف الگوریتم پیشنهادی شامل مراحل روش تشخیص چهره، الگوریتم یادگیری عمیق MTTS-CAN و تخمین ضربان قلب از سیگنال rPPG استخراج شده و معرفی معیارهای ارزیابی پرداخته شده است. در بخش سوم چگونگی پیاده‌سازی الگوریتم پیشنهادی بر روی دو دیتاست و نتایج ارائه شده و در بخش پایانی نیز بحث و جمع‌بندی انجام شده است.

روش‌ها

روش پیشنهادی برای تخمین ضربان قلب در این مطالعه شامل دریافت و آماده‌سازی دیتاست‌های عمومی و بومی، پردازش سیگنال PPG تماسی مربوطه و استخراج HR می‌باشد. سپس فریم‌های ویدئویی پیش‌پردازش شده و الگوریتم تشخیص چهره بر روی هر فریم اعمال شده است. پس از ساختاردهی شبکه عمیق این فریم‌های ویدئویی در مرحله تست به ورودی شبکه آموزش دیده اعمال شده و سیگنال rPPG تخمین می‌شود. ضربان قلب با استفاده از روش تشخیص پیک محاسبه شده و با ضربان قلب استخراج شده از سیگنال مرجع با استفاده از معیارهای رایج مقایسه می‌شوند. در ادامه این بخش هر یک از مراحل پایپ لاین پیشنهادی شرح داده خواهد شد.

مجموعه دادگان

روش پیشنهاد شده در این مقاله بر روی یک دیتاست عمومی (PURE) و یک دیتاست منتشر نشده با نام Stroop (جمع‌آوری شده در پژوهشگاه توسعه فناوری‌های پیشرفته) (۱۰) پیاده شده است. در این بخش این دو مجموعه داده معرفی می‌شوند.

مجموعه داده PURE شامل ۱۰ نفر است که حرکات مختلف و کنترل شده سر را در مقابل دوربین انجام می‌دهند. در طول این حرکات تصاویر سر و همچنین پالس مرجع (PPG تماسی) ثبت شده است. ثبت ویدئو و سیگنال مرجع برای هر ۱۰ نفر (۸ مرد، ۲ زن) در ٦ حالت مختلف انجام شده و در مجموع برای هر کدام حدود یک دقیقه ویدئو در دسترس است. ویدیوها با دوربین eco274CVGE توسط SVS-Vistek GmbH با نرخ فریم ۳۰ هرتز با وضوح ٦٤٠*٤٨٠ پیکسل ضبط شده‌اند. داده‌های مرجع به طور موازی با استفاده از یک پالس‌اکسیمتر گیره انگشتی (pulox CMS50E) که امواج نرخ پالس و SpO2 را با نرخ نمونه‌برداری ٦٠ هرتز ثبت می‌کند، گرفته شده است. آزمودنی‌ها با فاصله متوسط ۱/۱ متر مقابل دوربین قرار گرفتند. شرایط روشنایی نور روز از طریق یک پنجره بزرگ رو به رو تامین شده و ابرها با گذشت زمان کمی شرایط نور را تغییر می‌دادند. شش حالت مختلف این دیتاست شامل؛ ثابت بدون حرکت سر، صحبت کردن، حرکت آهسته سر، حرکت سریع سر و چرخش آهسته و متوسط سر می‌باشد.

مجموعه داده استروپ (۱۰) نیز از ۹۸ شرکت‌کننده ثبت شده است که هر ثبت در حدود ٦٠٠ ثانیه می‌باشد. در این دادگان، تصاویر مرئی توسط دوربین کینکت و با مشخصات نرخ فریم ۳۰ هرتز و وضوح ٦٤٠* ٤٨٠ ثبت شده است؛ در کنار تصاویر مرئی، سیگنال فوتوپلتیسموگرافی انگشتی به عنوان سیگنال مرجع با فرکانس نمونه‌برداری ٢٥٦ هرتز نیز دریافت شده است. فرد در طول ثبت تقریباً بدون حرکت در مقابل دوربین با فاصله حدود یک متر نشسته و تست استروپ را با ۵ سطح سختی انجام داده است.

تشخیص چهره

روش یادگیری عمیق MTTS-CAN پیشنهادی در (٢٤) شامل مرحله تشخیص چهره یا انتخاب ROI نیست. اما از آنجا که تصاویر

دوربین ها در شرایط واقعی تنها شامل سوژه مورد نظر نیست و فضای اطراف سوژه، وسایل یا افراد دیگر نیز در تصویر هستند (مشابه دیتاست‌های مورد استفاده در این پیاده‌سازی)، در نتیجه در این مقاله استفاده از روش تشخیص چهره به عنوان یک گام پیش پردازش پیشنهاد و اثر بخشی آن ارزیابی شده است. الگوریتم پردازش چهره مورد استفاده Mediapipe است. Mediapipe یک چارچوب Open source و چند پلتفرمی برای ساخت پایپلاین یادگیری ماشینی است که داده‌های صوتی و تصویری را پردازش می‌کند. مدل تشخیص چهره در Mediapipe به گونه‌ای طراحی شده است که در طیف وسیعی از شرایط نوری، جهت‌گیری‌ها و انسدادها به خوبی کار کند. همچنین برای عملکرد بلادرنگ بهینه شده است. شکل ۱، ورودی‌های شبکه یادگیری عمیق MTTS-CAN را در دو حالت با استفاده از تشخیص چهره و بدون استفاده از تشخیص چهره نشان می‌دهد. با توجه به اینکه در مدل MTTS-CAN فریم‌ها به ابعاد ۳۶*۳۶ تغییر سایز داده می‌شوند تا حد زیادی از کیفیت ورودی کاسته شده که مسلماً در نتیجه نهایی تخمین ضربان قلب موثر خواهد بود. همچنین میانگین درصد تشخیص چهره در ۹۸ آزمودنی مجموعه داده STROOP برابر ۷۹.۵۳ درصد می‌باشد. در صورت عدم تشخیص چهره فرد توسط الگوریتم تشخیص چهره Mediapipe در هر فریم، آن فریم به طور کامل به عنوان ورودی به شبکه عمیق وارد خواهد شد.

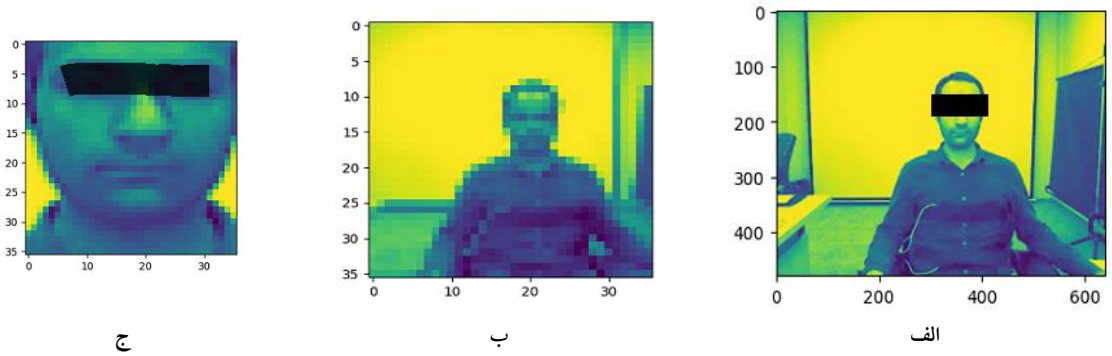
پیش‌پردازش و ساختاردهی شبکه

روش (۹) MTTS-CAN یک بهبود ساخته شده بر روی شبکه عمیق DeepPhys است. MTTS-CAN اطلاعات زمانی را از طریق معرفی یک ماژول شیفت زمانی (TSM) دریافت می‌کند. در روش MTTN-CAN تنها ۱۰ فریم برای تخمین ضربان قلب نیاز است که پتانسیل آن را برای استفاده در کاربردهای بلادرنگ نشان می‌دهد. این روش شامل دو شاخه موازی ظاهری (Appearance) و حرکتی (Motion) است. هر شاخه شامل چهار لایه کانولوشنی است. یک لایه‌ی اولیه درونی و یک لایه‌ی dropout به ترتیب بعد از دومین و چهارمین لایه کانولوشنی در نظر گرفته شده است. برای پیش‌پردازش ورودی شاخه‌ی ظاهری، هر فریم c(t) به اندازه ۳۶×۳۶ کاهش داده می‌شود که تعادلی بین حفظ کیفیت و حذف نویز دوربین ایجاد می‌کند. برای شاخه‌ی حرکتی، با استفاده از هر دو فریم مجاور به‌صورت $(c(t+1)-c(t))/(c(t)+c(t+1))$، فریم‌های نرمال شده محاسبه می‌شوند. فریم‌های نرمال شده در برابر تغییرات روشنایی کمتر حساس هستند و نسبت به فریم‌های خام c(t)، احتمال بیش‌برازش به مجموعه داده‌های خاص را کاهش می‌دهند (۹). پارامترهای کلیدی شبکه شامل لایه‌های کانولوشنی دو بعدی با اندازه کرنل ۳×۳، لایه‌های پولینگ متوسط با اندازه ۲×۲، و نرخ Dropout به ترتیب ۰/۲۵ و ۰/۵ برای جلوگیری از بیش‌برازش هستند. برای یادگیری شبکه، از بهینه‌ساز Adadelta با نرخ یادگیری ۱/۰ و اندازه دسته ۳۲ استفاده شده است.

شبکه در اصل در دو مرحله آموزش دیده است: ۱۲ دوره برای تنفس و ۲۴ دوره برای ضربان قلب. همچنین، از یک تابع زیان چندوظیفه‌ای (Multi-Task Loss) با وزن مساوی ($\alpha = 0/5$) برای بهینه‌سازی همزمان پیش‌بینی هر دو سیگنال استفاده شده است. البته در این مقاله تنها ضربان قلب با استفاده از این شبکه‌ی عمیق، تخمین زده شده است. وزن‌های شبکه آموزش دیده شده به صورت عمومی در دسترس است و نیاز به آموزش مجدد آن وجود ندارد. این شبکه روی دیتاست AFRL آموزش دیده است. در نتیجه مراحل آموزش و تست مدل در این پیاده‌سازی وجود ندارد.

پیشبینی

گام پیش‌بینی شامل تخمین سیگنال فوتوپلتیسموگرافی غیرتماسی (rPPG) از فریم‌های ویدئو ورودی با استفاده از مدل آموزش دیده (MTTS-CAN) است. در خروجی شبکه MTTS-CAN در مرحله‌ی پیش‌بینی سیگنال rPPG به دست می‌آید، در نتیجه در گام بعدی نیاز است که از سیگنال تخمین زده شده ضربان قلب استخراج شود.



شکل ۱. الف) یک کانال ازتصویر فریم ورودی (۴۶۰*۶۸۰)، ب) فریم پیش‌پردازش شده (۳۶*۳۶)، ج) فریم پیش‌پردازش شده با استفاده از تشخیص چهره (۳۶*۳۶)

پس‌پردازش

پس‌پردازش بر روی سیگنال rPPG، شامل Detrend کردن سیگنال و فیلتر کردن آن می‌باشد. Detrend می‌تواند در بهبود دقت پیش‌بینی با حذف الگوهای سیستماتیک در داده‌ها که با دینامیک اصلی سیستم مرتبط نیستند، مورد استفاده قرار گیرد. فیلتر کردن نیز اجزایی غیر فیزیولوژیک یا غیر مرتبط با سیگنال PPG را حذف می‌کند. بنابراین سیگنال با یک فیلتر باتروورث میان‌گذر با فرکانس قطع ۰/۷۵ تا ۲/۵ هرتز فیلتر شد.

ضربان قلب با میانگین‌گیری تمام فواصل میان پیک‌ها IBI (Inter beat Interval) در یک پنجره زمانی محاسبه و سپس معکوس می‌شود: یعنی HRw = 1/IBIw، که در آن IBIw میانگین تمام فواصل بین ضربانی است که در پنجره زمانی w قرار می‌گیرند. این فرمول ضربان قلب را بر حسب هرتز ارائه می‌دهد (با فرض IBI برحسب ثانیه) و در صورتی که آن را در ۶۰ ضرب کنیم، ضربان قلب را بر حسب ضربان در دقیقه (bpm) خواهد بود. انتخاب پنجره زمانی (w) می‌تواند بر اساس نیاز کاربر باشد (به عنوان مثال، HR لحظه‌ای، HR طولانی‌مدت) (۱۱). در بخش بحث چگونگی انتخاب طول مناسب پنجره تخمین ضربان قلب توضیح داده شده است.

معیارهای ارزیابی

معیارهای RMSE (Root Mean Squared Error) و MAE (Mean Absolute Error) برای اندازه‌گیری خطا و دقت الگوریتم‌های پیش‌بینی استفاده می‌شوند. هر دو معیار برای اندازه‌گیری خطا در پیش‌بینی‌های عددی به کار می‌روند. در مقالات تخمین ضربان قلب این دو معیار برای اندازه‌گیری خطای تخمین ضربان قلب بر حسب (bpm) رایج هستند.

MAE میانگین اختلاف مطلق بین مقادیر واقعی و پیش‌بینی شده است. این معیار مجموع خطاهای مطلق را بر تعداد نمونه‌ها تقسیم می‌کند (رابطه ۱). RMSE، میانگین مجذور خطا را نشان می‌دهد و معیار مناسبی برای محاسبه خطاهای بزرگ است (رابطه ۲). در این روابط، n تعداد نمونه‌هاست، yi مقدار واقعی و ŷi مقدار پیش‌بینی شده توسط الگوریتم است.

$$MAE = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j| \qquad \text{(رابطه ۱)}$$

$$RMSE = \sqrt{\sum_{i=1}^{n}\frac{(\hat{y}_i - y_i)^2}{n}} \qquad \text{(رابطه ۲)}$$

یافته‌ها

در این مقاله، روش یادگیری عمیق MTTS-CAN همراه با الگوریتم تشخیص چهره mediapipe به منظور تخمین ضربان قلب به صورت غیر تماسی با استفاده از تصاویر ویدئو مرئی پیاده‌سازی شده است. به منظور کاهش خطا، سیگنال rPPG در پنجره های ۶۰ ثانیه‌ای تخمین زده شده و در نهایت با استفاده از روش یادگیری عمیق مذکور به ازای فریم‌های ویدئوی ورودی، سمپل‌های سیگنال rPPG تخمین زده شد. پس از آن با یک گام پس پردازش، ضربان قلب در پنجره‌های ۲۰ ثانیه‌ای با همپوشانی ۹۵ درصد (به ازای هر یک ثانیه یک مقدار برای ضربان قلب) به دست آمد. با پیاده‌سازی این گام‌ها برای هر دو دیتاست PURE و استروپ مقادیر ضربان قلب محاسبه شدند. مقادیر تخمین زده شده با مقادیر ضربان قلب مرجع هر دیتاست مقایسه شده و نتایج آن در این بخش ارائه شده است.

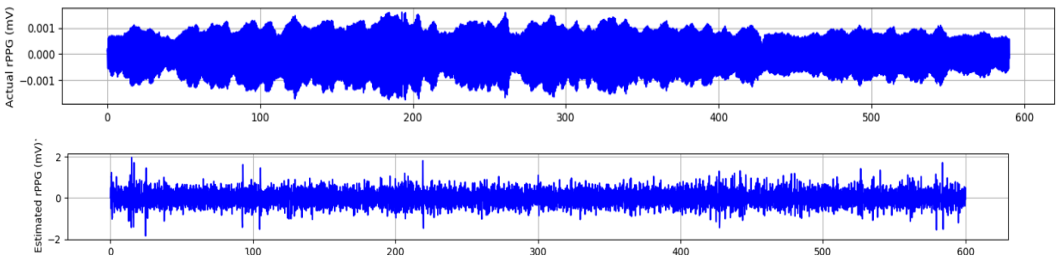
نتایج پیاده‌سازی الگوریتم پیشنهادی بر روی دیتاست‌های PURE و استروپ برای دو حالت استفاده از روش پیش‌پردازش تشخیص چهره و بدون استفاده از تشخیص چهره در جدول ۱ ارائه شده است. همچنین نمونه‌ای از سیگنال rPPG تخمین زده شده و سیگنال PPG تماسی مربوطه در شکل ۲ نمایش داده شده‌اند.

جدول ۱. مقایسه‌ی میانگین خطای تخمین ضربان قلب در دیتاست PURE و STROOP با روش‌های پیشنهادی

میانگین	تشخیص چهره ندارد		تشخیص چهره	
مجموعه داده‌ها	PURE	STROOP	PURE	STROOP
MAE bpm	۷/۹۹	۶/۳۶	۷/۷۲	۵/۳۷
RMSE bpm	۸/۷۳	۷/۴۲	۸/۴۴	۶/۵۱

الف

ب



شکل ۲. الف) سیگنال PPG تماسی ground truth و ب) سیگنال rPPG تخمین زده شده یک سوژه از دیتابیس STROOP

بحث

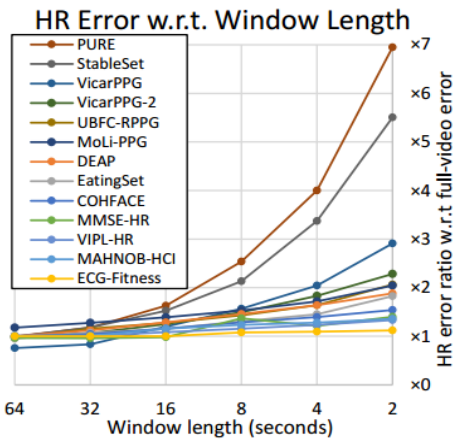
انتخاب طول پنجره‌ای که ضربان قلب بر روی آن محاسبه می‌شود، بسیار مهم است. تخمین ضربان قلب در پنجره‌های زمانی بزرگتر کار ساده‌تری نسبت به تخمین ضربان قلب آنی در پنجره‌های کوتاه‌تر است. این امر به این دلیل است که به طور میانگین تغییرات در فواصل بین ضربان، ناشی از پیک‌های نادرست شناسایی شده و از دست رفته در طول یک پنجره بزرگ، کمتر است.

شکل ۳، نرخ خطا را هنگام استفاده از طول پنجرهای مختلف با توجه به میانگین بلندمدت خطای ضربان قلب در کل ویدیو بر روی دیتاست‌های رایج نشان می‌دهد. به وضوح می‌توان مشاهده کرد که نرخ خطا برای اکثر مجموعه داده‌ها با کاهش طول پنجره به صورت تصاعدی افزایش می‌یابد. مجموعه داده‌هایی که کمتر تحت تأثیر طول پنجره قرار می‌گیرند، مواردی هستند که خطاهای ناشی از عوامل دیگر (مانند حرکت، فشرده‌سازی و غیره) تأثیر طول پنجره را تحت‌الشعاع قرار می‌دهند (به عنوان مثال، ECG-Fitness، MAHNOB-HCI). این آزمایش نشان می‌دهد که اندازه پنجره انتخابی می‌تواند بر دقت تخمین ضربان قلب تأثیر بگذارد و باید در طول ارزیابی‌ها در نظر گرفته شود (۲۵). با توجه به این مهم، برای پیاده‌سازی در این پروژه طول پنجره تخمین ضربان قلب از ویدئو مرئی ۶۰ ثانیه در نظر گرفته شد.

جدول ۲. مقایسه‌ی میانگین خطای روش‌های سنتی و یادگیری عمیق در تخمین ضربان قلب غیرتماسی با روش پیشنهادی مقاله بر روی دیتاست عمومی PURE

روش‌ها/معیارها	RMSE bpm	MAE bpm
CHROM (12)	۱۱/۳۶	۶/۲۹
HR-CNN (13)	۱۱/۰۰	۸/۷۲
MTTS-CAN (14)	۸/۴۵	-
پیشنهادی	۸/۴۴	۷/۷۲

بر اساس نتایج به‌دست آمده در این پژوهش، روش یادگیری عمیق MTTS-CAN برای تخمین ضربان قلب از تصاویر ویدئویی، کارآمد است. روش پیشنهادی در این مقاله قابلیت استفاده در سامانه‌های برخط تشخیص ضربان قلب را دارا می‌باشد. مشکلات مربوط به حافظه و سرعت که اغلب در هنگام استفاده از روش‌های یادگیری عمیق غیرقابل اجتناب هستند، در رویکرد پیشنهادی مقاله حل شده و دقت تخمین قابل قبولی برای ضربان قلب غیرتماسی ارائه شده است. علاوه بر آن، نتایج نشان می‌دهند که روش پیشنهادی به دیتاست خاصی وابسته نبوده و در یک دیتاست بومی در مقایسه با دیتاست بنچمارک نتایج مناسبی ایجاد کرده است. جدول ۲، نتایج تخمین ضربان قلب با روش‌های سنتی و عمومی به دست آمده در مقاله‌ی حاضر را با نتایج پیاده‌سازی روش MTTS-CAN بر روی دیتاست PURE مقایسه می‌کند.



شکل ۳. نسبت خطای تخمین ضربان قلب (MAE) با طول پنجره زمانی متغیر با توجه به میانگین خطای تخمین ضربان قلب در درازمدت در ویدیوی کامل (۱۱)

نتیجه‌گیری

در این مقاله ما یک پایپ لاین استخراج ضربان قلب غیرتماسی از تصاویر دوربین مرئی پیشنهاد دادیم. با استفاده از یک روش تشخیص چهره جدید و یک الگوریتم یادگیری عمیق آموزش داده شده، سیگنال فوتوپلتیسموگرافی غیر تماسی از فریم‌های ویدئویی تخمین زده شد. پس از آن ضربان قلب از این سیگنال‌ها با دقت قابل قبول محاسبه شد. نتایج گزارش شده بر روی دو دیتاست PURE و استروپ، کارآمدی روش پیشنهادی را تأیید کردند.

تشکر و قدردانی

نویسندگان این مقاله از اعضای وقت گروه پردازش پژوهشگاه توسعه فناوری‌های پیشرفته که دیتاست استروپ با تلاش آنها تهیه و آماده استفاده شد، تشکر و قدردانی می‌گردد..

References

1. Leonhardt S, Leicht L, Teichmann D. Unobtrusive vital sign monitoring in automotive environments—a review. Sensors (Basel) 2018; 18(9): 3080.
2. Lorato I, Stuijk S, Meftah M, Kommers D, Andriessen P, van Pul C, et al. Multi-camera infrared thermography for infant respiration monitoring. Biomed Opt Express 2020; 11(9): 4848-61.
3. Chan P, Wong G, Dinh Nguyen T, Nguyen T, McNeil J, Hopper I. Estimation of respiratory rate using infrared video in an inpatient population: an

observational study. J Clin Monit Comput 2020; 34(6): 1275–84.
4. Spicher N, Kukuk M, Maderwald S, Ladd ME. Initial evaluation of prospective cardiac triggering using photoplethysmography signals recorded with a video camera compared to pulse oximetry and electrocardiography at 7T MRI. BioMed Eng OnLine 2016; 15(1): 126.
5. Wang J, Spicher N, Warnecke JM, Haghi M, Schwartze J, Deserno TM. Unobtrusive health monitoring in private spaces: the smart home. Sensors (Basel) 2021; 21(3): 864.
6. Wang J, Warnecke J, Haghi M, Deserno T. Unobtrusive health monitoring in private spaces: the smart vehicle. Sensors (Basel) 2020; 20(9): 2442.
7. Selvaraju V, Spicher N, Wang J, Ganapathy N, Warnecke JM, Leonhardt S, et al. Continuous monitoring of vital signs using cameras: a systematic review. Sensors (Basel) 2022; 22(11): 4097.
8. Hirzi MF, Efendi S, Sembiring RW. Literature study of face recognition using the viola-jones algorithm. in: 2021 international conference on Artificial Intelligence and Mechatronics Systems (AIMS) Bandung, Indonesia: IEEE; 2021 [cited 2024 Mar 28]. Available from: https://ieeexplore.ieee.org/document/9466010/
9. Liu X, Fromm J, Patel S, McDuff D. Multi-Task Temporal Shift Attention Networks for On-Device Contactless Vitals Measurement.
10. Shahi K, kaveh R,Saeedi M, Godarzi MM, Babaee N, Rezaee M, Gholi pour Hasan, Azizi A, Pejman H. Multimodal data (physiological signals and Thermal images) from individual to provide a continous index of different levels of mental arousal. Proceedings of the 26th Nathonal Conference and 4th International Conference on Biomedical Enginnering of Tehran, Iran. 2019.[in Persian]
11. Gudi A, Bittner M, Van Gemert J. Real-time webcam heart-rate and variability estimation with clean ground truth for evaluation. Appl Sci 2020; 10(23): 8630.
12. de Haan G, Jeanne V. Robust pulse rate from chrominancebased rppg. IEEE Transactions on Biomedical Engineering 2013; 60(10): 2878–86.
13. Špetlík R, Vojtech F, Jirí M. Visual heart rate estimation with convolutional neural network. Proceedings of the British machine Vision Conference, Newcastle, UK, 2018. p. 3-6.
14. Deniz Cyli N. rPPG based heart rate estimation using deep learning [Thesis]. Istanbul: Marmara University; 2021. .

DOI: *10.48305/jims.v43.i836.1383*

Journal of Isfahan Medical School **Vol. 43, No. 836, 4th Week, December 2025**

Received: 09.12.2024 Accepted: 22.11.2025 Published: 20.12.2025

Remote Heart Rate Estimation Based on Deep Learning

Sara Pourmohammadi[1], Kian Shahi[2]

Original Article

Abstract

Background: Using remote photoplethysmography (rPPG) technology, monitoring heart rate (HR) has become possible without physical contact. Over the past decades, methods have been developed to estimate rPPG signals and heart rates using video frames. Recently, deep learning techniques have also been applied in this field, showing promising performance.

Methods: In this study, a deep neural network (MTTS-CAN) along with a face detection algorithm (MediaPipe) was used to estimate heart rate remotely from videos of individuals in a public dataset (PURE) and a local dataset (Stroop).

Findings: The implementation results on the PURE dataset are comparable to those published in articles (MAE: 7.72 bpm). Moreover, the results on the local dataset are also acceptable (MAE: 5.53 bpm).

Conclusion: This paper presents an acceptable accuracy for non-contact heart rate estimation. Additionally, the results indicate that the proposed method is not dependent on a specific dataset and has produced satisfactory results in a local dataset compared to the benchmark dataset.

Keywords: Heart Rate; Remote Heart Rate Monitoring; rPPG Signal; PURE Dataset; Deep Learning

Citation: Pourmohammadi S, Shahi K. **Remote Heart Rate Estimation Based on Deep Learning.** J Isfahan Med Sch 2025; 43(836): 1383- 9.

1- PhD, Signal Processing Group, Research Center for Development of Advanced Technologies, Tehran, Iran
2- MSc, Signal Processing Group, Research Center for Development of Advanced Technologies, Tehran, Iran
Corresponding Author: Sara Pourmohammadi, PhD, Signal Processing Group, Research Center for Development of Advanced Technologies, Tehran, Iran; Email: Pourmohammadi@semnan.ac.ir