بسمه تعالی

دانشکده برق و کامپیوتر

گزارش پروژه درس شبکه‌های عصبی

آشکار سازی چهره بر مبنای شبکه‌های عصبی

استاد: دکتر ترکمنی آذر

تهیه کننده: علی بهلولی

زمستان ۱۳۸۱

چکیده:

در این مقاله سیستمی برای تشخیص صورت افراد ارایه خواهد شد. شبکه عصبی یک پنجره کوچکی از تصویر را دریافت می کند و تصمیم می گیرد که آیا این قسمت از تصویر بخشی از صورت است یا خیر.برای اینکه کارایی شبکه از یک شبکه تک بیشتر شود، سیتم نتیجه چندین شبکه را مقایسه می کند ودر نهایت تصمیم خود را اعلام می کند.

برای Train شبکه از یک الگوریتم خود راه‌انداز[1] استفاده می‌شود، به طوری که تعدادی تشخیص نادرست جزء الگوها ی آموزشی گذاشته می‌شود. بدست آوردن الگوهای آموزشی به صورت دستی کار مشکلی است ، چون باید باید تمامی قسمتهای تصویر را بررسی کرد.در مقایسه بقیه روشهای تشخیص صورت، این سیستم در تشخیص و رد کردن نمونه‌ها دارای سرعت بیشتری است.

[1] bootstrap

۱ – مقدمه:

در این مقاله روشی مبتنی بر شبکه‌های عصبی برای تشخیص قسمت جلو صورت در تصاویر غیر رنگی ارایه خواهد شد.

الگوریتم و همجنین روش یادگیری به صورت کلی ارایه شده است به نحوی که علاوه براین که از این روش برای تشخیص بقیه قسمتهای صورت نیز می‌توان استفاده کرد همجنین برای کارهای مشابه که نیاز به شناسایی الگو دارد نیز میتواند استفاده گردد.

آموزش شبکه عصبی برای تشخیص صورت، به خاطر مشکل بودن تشخیص تصاویر غیر صورت، همواره مانند مسابقه‌ای بوده که افراد زیادی در این مورد کار کرده‌اند.

برخلاف Face rcognition که برای جدا کردن تصاویر صورت مختلف به کار می‌رود، درعمل face detection باید تصاویری که دارای صورت هستند از تصاویر دیگر جدا شوند.

ارایه مثال برای تصاویری که صورت در آنها باشد بسیار آسان است ولی ارایه مثال برای تصاویری که صورت در آنها نیست ، بسیار مشکل است.مجموعه الگوهای آموزشی برای دسته دوم دارای رشد بسیار کمی هستند.

در طراحی این سیستم سعی شده است به جای زیاد کردن مجموعه الگوهای آموزشی بدون صورت، تعداد تصاویر بیشتری در مرحله آموزش اعمال شوند.

در قسمت دوم در مورد این روش آموزش بیشتر شرح داده خواهد شد. در قسمت سوم کارایی این سیستم آزمایش می‌شود. نشان داده خواهد شد که این سیستم دارای توانایی تشخیص 90.5% صورت در ۱۳۰ الگوی آزمایش با تعداد قابل قبولی خطا در تشخیص است. در قسمت چهارم این سیستم با سیستم مشابه مقایسه خواهد شد. در فصل پنجم نتیجه گیری و پیشنهاداتی برای ادامه کار آورده خواهد شد.

۲ – توصیف سیستم

این سیستم دارای ۲ مرحله است : در مرحله اول مجموعه‌ای از فیلترهای مبتنی بر شبکه‌های عصبی به تصویر اعمال می‌شوند و سپس خروجی این فیلترها به داوری گذاشته می‌شوند.

فیلترها در مقیاسهای متفاوت قسمتهای مختلف تصویر را آزمایش می کنند و به جستجوی قسمتهایی که احتمال وجود صورت هستند، می‌پردازند. در نهایت قسمتهایی که مشترکا توسط فیلترهای مختلف صورت تشخیص داده شده‌اند به عنوان نتیجه اعلام خواهند شد.

۲-۱ مرحله اول: فیلترهای مبتنی بر شبکه عصبی

اولین قسمت از سیستم فیلتری است که دارای ورودی ناحیه‌ای 20x20 از تصویر است و خروجی آن عددی است بین 1− تا1 ، که نشان دهنده وجود یا عدم وجود صورت است.برای تشخیص صورت در قسمتهای دیگر این فیلتر به تمامی قسمتهای تصویر اعمال می‌شود.برای تشخیص صورتهایی که از 20x20 بزرگترند ، تصویر ورودی با فاکتور ۱،۲ به صورت مکرر نمونه برداری می شود و فیلتر به همه این مقیاسها اعمال خواهد شد.

در شکل ۱ الگوریتم اعمال فیلتر نشان داده شده است. در اولین مرحله پردازش قسمت وفقی به قسمتی از تصویر اعمال می‌شود. سپس این قسمت از شبکه عصبی عبور می کند، در این مرحله تصمیم گیری می‌شود که آیا تصویر صورت است یا نه؟. پیش پردازش اولیه سعی می کند شدت روشنایی را در اطراف تصویر یکسان کند. تابعی که به صورت خطی د رعرض پنجره وارد شده حرکت می کند و شدت نور را یکسان می کند، بنابرین این شدت روشنایی در محاسبه‌ها صرف نظر می‌شود.

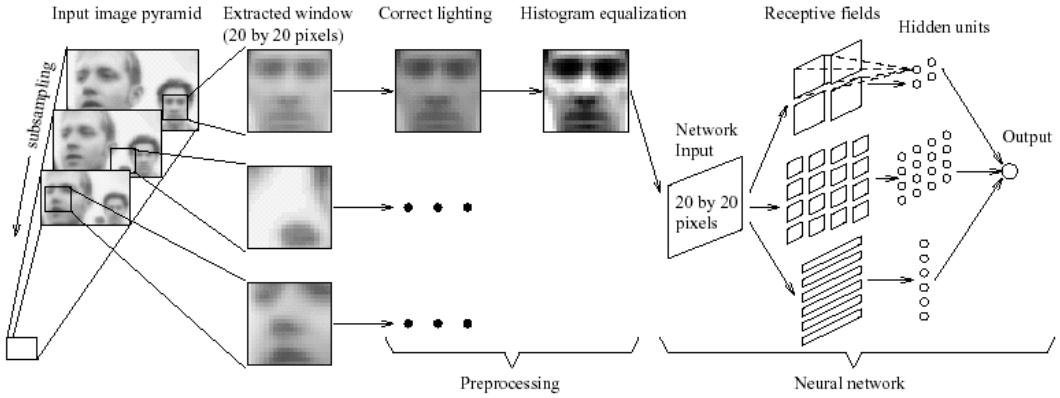


Figure 1: The basic algorithm used for face detection.

این تابع خطی مقدار روشنایی کلی در این قسمت از تصویر را به صورت تقریبی محاسبه کرده و از همین قسمت تفریق می کند. محلهای تغییر روشنایی را جبران کند.سپس برای پخش کردن شدت روشنایی در تمام تصویر، عمل .histogram equ انجام می‌شود. این هیستوگرام برای هر پیکسل درون پنجره محاسبه می‌شود. این عمل باعث جبران تفاوتهایی که به خاطر دوربین عکاسی در تصاویر مختلف ایجاد می‌شود، می گردد. همچنین کنتراست را در بعضی شرایط بهتر می کند.هر پنجره بعد از اینکه ای عملیات پرادازشی اولیه روی آن صورت گرفت به شبکه عصبی فرستاده می‌شود.این شبکه در لایه ورودی دارای اتصالات شبکیه‌ای[۲] است. فیلدهای گیرنده در واحدهای مخفی در شکل ۱ نشان داده شده است.

[2] retinal

سیستم دارای ۳ نوع واحد مخفی می‌باشد که عبارتند از :

۱- واحد مخفی ۴ تایی که به زیرناحیه‌ها به صورت 10x10 نگاه می‌کند.

۲- واحد مخفی ۱۶ تایی که به زیرناحیه‌ها به صورت 5x5 نگاه می‌کند.

۳- واحد مخفی ۴ تایی که به زیرناحیه‌ها به صورت 20x5 و به صورت افقی ومایل نگاه می‌کند.

هر کدام از این نوعها به واحد مخفی اجازه می‌دهند که نمایشهای محلی که شاید اطلاعاتی در مورد صورت داشته باشند را بدست آورند، علاوه بر این در تصویر یک لایه تک از واحد مخفی برای هر ورودی نمایش داده شده است، این واحدها می‌توانند تکرار شوند.

برای آزمایش که انجام شده است و در ادامه شرح داده می‌شود، شبکه‌ای با ۲ و۳ مجموعه از این واحدهای مخفی استفاده شده است.(مشابه کارهایی که در تشخیص کاراکتر و صوت به کار می‌رود).شبکه دارای تنها یک خروجی است که نشان می‌دهد آیا پنجره شامل صورت بوده یا نه؟

برای آموزش شبکه عصبی که در مرحله اول برای ساختن فیلتری دقیق استفاده شده ، به تعداد زیادی تصویر صورت و تصاویری که شامل صورت نیستند ، نیاز دارد.در حدود ۱۰۵۰ تصویر صورت از دیتابیس CMU در دانشگاه هاروارد استفاده شده است. این تصویرها شامل تصاویر صورت در اندازه‌های مختلف و در جهتهای مختلف و با مکانها و نورهای متفاوت است. چشمها ولبه‌های بالایی آن به صورت دستی مشخص شده‌اند.این نقطه‌ها باعث شده که تصاویر گوناگون در سایزهای و جهتها و مکانهای متفاوت به صورت زیر نرمالیزه شوند:

۱- چرخش تصویر به گونه‌ای که دو چشم در یک خط افقی قرار بگیرند.

۲- تغییر مقیاس تصویر به گونه‌ای که فاصله بین دو چشم تا لبه بالیی تصویر برابر ۱۲ پیکسل شود.

۳- یک پیکسل بالای چشم را به یک ناحیه 20x20 گسترش داده می‌شود.

در مجموعه الگوهای آموزشی از هر تصویر اصلی با چرخش تصادفی(حداکثر ۱۰ درجه)، ۱۵ صورت ایجاد می‌شود . سپس به هر پنجره 20x20 در این مجموعه ،پردازشهای اولیه اعمال می‌گردند. تصادفی کردن باعث می‌شود که تغییرات فیلتر اعمال شده کم و در حدود 10% باشد. تغییرات بیشتر در عمل تبدیل و مقیاس بستگی به این دارد که که فیلتر به چه نحوی اعمال گردد. به عبارت دیگر با ضریب 1.2 مقیاس انجام شده باشد. برای تمرین می‌توان هر تصویری را به عنوان تصویر غیر صورت‌به شبکه اعمال کرد به علت اینکه فاصله تصاویر غیر صورت خیلی بیشتر از تصاویر صورت است.با وجود این ایجاد دسته‌ای از تصاویر غیر صورت کار مشکلی است.

به جای اینکه الگوهای غیر صورت را قبل از شروع عملیات یادگیری آماده شوند می‌توان به صورت زیر عمل کرد:

۱- ایجاد مجموعه‌ای از تصاویر غیر صورت با تولید ۱۰۰۰ تصویر با پیکسل‌های تصادفی و شدت نور متفاوت.و عملیات پردازشی اولیه روی تک تک این تصاویر اعمال شوند.

۲- آموزش شبکه عصبی برای تولید ۱ برای تصاویر صورت و ۱- برای نمونه‌های غیر صورت.الگوریتم یادگیری بر پایه برگشت خطا[3] است.در اولین ایتریشن این حلقه وزنهای شبکه به صورت تصادفی انتخاب می‌شوند. در مراحل بعدی وزنها بر اساس آموزش قبلی اصلاح می‌شوند.

۳- یک تصویر غیر صورت به سیستم اعمال می گردد.مجموعه زیربخشها به صورت نادرست آن را تصویر صورت اعلام می کند.

٤- حدود ۲۵۰ عدد از این زیربخشها را به صورت تصادفی به مرحله پیش پردازش اعمال می‌شوند. و آنها به الگوهای اشتباه اضافه می‌شوند. سپس به مرحله ۲ برمی گردیم.

برای مرحله راه‌اندازی از ۱۲۰ تصویر منظره برای تصاویر غیر صورت استفاده شده است. یک آموزش نوعی شامل ۸۰۰۰ تصویر غیر صورت از ۱۴۶۲۱۲۱۷۸ زیر بخش است.

۲-۲ مرحله دوم: جمع آوری پاسخهای مشابه و داوری

سیستمی که شرح آن داده شد از یک شبکه عصبی استفاده میکند و احتمال تشخیص اشتباه در آن وجود دارد. در ادامه به روشهایی برای کاهش خطا با جزییات بیتر آورده می‌شود

به علت مکانهای کوچک و غیر متغیر در فیلتر ، تصاویر واقعی اغلب در اطراف صورت به صورت چند گانه تشخیص داده می‌شوند(چندین تصویر صورت به جای یکی) در حالی که تشخیصهای اشتباه در مکانهای تک رخ می‌دهند. با گذاشتن یک حد مینیمم در تعداد تشخیص‌ها از تعداد زیادی از خطاها می‌توان جلوگیری کرد. روش ابتکاری دیگر بر این اساس است که صورت معمولا دارای همپوشانی خیلی کم در تصویر است. بنابرین اگر ۲ تشخیص با یکدیگر همپوشانی داشتند می‌توان تشخیصی را که دارای احتمال کمتری‌است را حذف کرد.

در زمان آموزش، سلسله مراتب شبکه با وزنهای متفاوت مجموعه‌های متفاوت از تصاویر غیر صورت انتخاب می کنند. تولید بایاسهای متفاوت باعث ایجاد اشتباه‌های متفاوت می‌شود. به کمک رای گیری از شبکه‌های متفاوت می‌توان نتیجه بهتری گرفت. به عنوان مثال اگر دو واحد اعلام تشخیص کردند خروجی یک شود.

[3] error backpropagation

۳- نتایج آزمایشهای انجام شده:

این سیستم توسط سه دسته بسیار بزرگ از تصاویر تست شد. که به صورت کامل با الگوهای آموزشی متفاوت بودند.

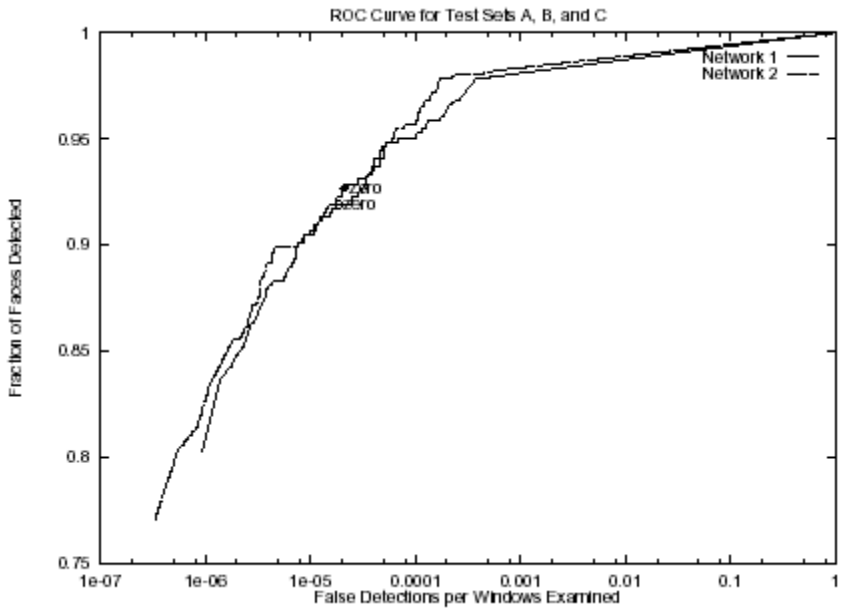
مجموعه A در CMU جمع آوری شده بود و شامل ۴۲ تصویر اسکن شده و تصاویر روزنامه و تصاویری که از روی web جمع‌آوری شده بود و تصاویر باینری شده تلوزیون بود. در این تصاویر ۱۶۹ صورت وجود داشت و نیاز بود که شبکه، ۲۲۰۵۳۱۲۴ پنجره 20x20 پیکسلی را آزمایش کند.

مجموعه B شامل ۲۳ تصویر بود که دارای ۱۵۵ تصویر صورت بود.(۹۷۶۸۰۸۴ پنجره)

مجموعه C شبیه مجموعه A بود ولی شامل تصاویر با پیچیدگی بیشتر و بدون چهره بودند تا دقت اندازه‌گیر تشخیص‌های غلط بالا رود. این مجموعه شامل ۶۵ تصویر با ۱۸۳ چهره و ۵۱۳۸۰۰۳ پنجره بود.

علاوه بر خروجی باینری که تولید می‌شد خروجی فیلترها عددی بین ۱-و ۱ بود که تشان می‌داد چهره‌ای موجود است یا خیر؟

موقع آموزش هنگام انتخاب الگوهایی که شامل چهره نبودند مقدار حد برابر صفر گرفته می‌شد.(اگر خروجی بزرگتر از صفر می‌شد یک اشتباه فرض می‌شد)علاوه براین هنگام آموزش این مقدار تغییر می‌کرد تا از اوضاع سیستم آگاه می شدیم.مقدار تشخیص را برای حدهای بین ۱- تا ۱ را اندازه‌گیری کردیم. زمانی که حد را برابر ۱ در نظر می‌گرفتیم احتمال تشخیص غلط صفر بود. ولی هیچگونه چهره‌ای شناسایی نمی‌شد. همانطور که مقدار حد کاهش می‌یافت تعداد تشخیصهای درست بالا می‌رفت ولی در عین حال تعدادی تشخیص نادرست صورت می‌گرفت.نمودار این tradeoff در شکل ۲ آمده است.در این شکل سرعت تشخیص بر حسب تعداد اشتباه‌های رخ داده و برای دو شبکه مستقل رسم شده است.



در جدول ۱ میزان کارآیی ۴ شبکه تنها آورده شده است. همپوشانی باعث شده است که چندین تشخیص در یک ناحیه رخ دهد. همچنین در این جدول نتیجه and,or کردن شبکه‌های مختلف‌و استفاده از یک رای گیر آورده شده است.

شبکه ۳ و ۴ زیرشاخه‌ای از شبکه‌های ۱و۲ هستند. نسبتا الگوهای اشتباه در ترتیبهای متفاوت در زمان آموزش شبکه به آن وارد شده‌اند.نتیجه and.or کردن شبکه ها بر پایه شبکه‌های ۱و۲ هستند. در این جدول درصد تشخیصهای درست آورده شده است همچنین تعداد تشخیصهای نادرست برای الگوهای تست A,B,Cنیر آورده شده است.

Table 1: Combined detection and error rates for Test Sets A, B, and C

Type	System	Missed faces	Detect rate	False detects	False detect rate
	0) Ideal System	0/507	100.0%	0	0/83099211
Single network, no heuristics	1) Network 1 (52 hidden units, 2905 connections)	37	92.7%	1768	1/47002
	2) Network 2 (78 hidden units, 4357 connections)	41	91.9%	1546	1/53751
	3) Network 3 (52 hidden units, 2905 connections)	44	91.3%	2176	1/38189
	4) Network 4 (78 hidden units, 4357 connections)	37	92.7%	2508	1/33134
Single network, with heuristics	5) Network 1 → threshold(2,1) → overlap elimination	46	90.9%	844	1/98459
	6) Network 2 → threshold(2,1) → overlap elimination	53	89.5%	719	1/115576
	7) Network 3 → threshold(2,1) → overlap elimination	53	89.5%	975	1/85230
	8) Network 4 → threshold(2,1) → overlap elimination	47	90.7%	1052	1/78992
Arbitrating among two networks	9) Networks 1 and 2 → AND(0)	66	87.0%	209	1/397604
	10) Networks 1 and 2 → AND(0) → threshold(2,3) → overlap elimination	107	78.9%	8	1/10387401
	11) Networks 1 and 2 → threshold(2,2) → overlap elimination → AND(2)	74	85.4%	63	1/1319035
	12) Networks 1 and 2 → thresh(2,2) → overlap → OR(2) → thresh(2,1) → overlap	48	90.5%	362	1/229556
Three nets	13) Networks 1, 2, 3 → voting(0) → overlap elimination	53	89.5%	195	1/426150

کارایی سیستم در حالتی که از چندین شبکه عصبی و رای گیر استفاده می‌شود خیلی بهتر از حالات دیگر است.

سیستم ۱ تا ۴ نشان دهنده کارایی خام شبکه هستند. سیستمهای ۵ تا۸ از همان شبکه ها استفاده می‌کنند ولی دارای مقدار حدی و همچنین مرحله همپوشانی هستند که باعث کاهش تعداد تشخیصهای اشتیاه می‌شوند. در عوض باعث کاهش سرعت تشخیص چهره می‌شود.تمامی سیستم‌های باقیمانده از رای‌گیر با شبکه‌ها ی متعدد استفاده می‌کنند. رای‌گیر باعث می‌شود که سرعت تشخیصهای غلط کاهش یابد.دقت کنید که در سیستم‌هایی که از رای‌گیر استفاده می‌کنند نسبت تشخیصهای غلط بی نهایت کوچک می شود. حدود ۱ خطا در ۲۲۹۵۵۶یا کمتر از آن حدود ۱ خطا در ۱۰۳۸۷۴۰۱ .این نسبت به نوع رای‌گیر نیز بستگی دارد.سیستمهای ۱۰و ۱۱ و ۱۲ نشان می‌دهد که تشخیص دهنده می‌تواند تا مقدار خیلی کوچکی پیش رود. سیستمم ۱۰ که از And استفاده می‌کند دارای کمترین مقدار تشخیص غلط و دارای سرعت تشخیص حدود ۷۹ درصد می‌باشد. در طرف دیگر سیستم ۱۲ که از Or استفاده می‌کند، دارای سرعت تشخیص بالاتری، حدود ۹۰ درصد ولی در عوض تعدادتشخیصهای غلط آن خیلی بیشتراز سیستم قبلی است. سیستم ۱۱ سیستمی بین این دو مورد است تفاوت کارایی این سیستم با سیستمهای دیکر را با استراتژی رای‌گیر آن می‌توان درک کرد. هنگامی که از And استفاده می‌شود یک تشخیص اشتباه در کمترین سرعت است در حالی که هنگامی که از or استفاده می‌شود تشخیص چهره بدرستی، توسط یک شبکه انجام می‌گیرد. سیستم ۱۳ که از رای ۳ شبکه استفاده می‌کند دارای سرعت تشخیص مشابه ولی تعداد اشتباههای کمتری است.

نتیجه‌ای که از این جدول گرفته می‌شود این است که سیستم ۱۱ دارای tradeoff معقولی بین تعداد تشخیصهای غلط و سرعت تشخیص درست می‌باشد. سیستم ۱۱ به صورت متوسط ۸۶درصد از چهره‌ها را بدرستی تشخیص می‌دهد. و دارای متوسط تشخیص غلط ۱ خطا در ۱۳۱۹۰۳۵ پنجره می‌باشد.در شکل ۳ مثالهایی از خروجی تصویر سیستم ۱۱ آورده شده است.

Table 2: Comparison of [Sung and Poggio, 1994] and our system on Test Set B

System	Missed faces	Detect rate	False detects	False detect rate
10) Networks 1 and 2 → AND(0) → threshold(2,3) → overlap elimination	34	78.1%	3	1/3226028
11) Networks 1 and 2 → threshold(2,2) → overlap elimination → AND(2)	20	87.1%	15	1/645206
12) Networks 1 and 2 → threshold(2,2) → overlap → OR(2) → threshold(2,1) → overlap	11	92.9%	64	1/151220
[Sung and Poggio, 1994] (Multi-layer network)	36	76.8%	5	1/1929655
[Sung and Poggio, 1994] (Perceptron)	28	81.9%	13	1/742175

۴- مراجع

1) [Hunke, 1994] H. Martin Hunke. Locating and tracking of human faces with neural networks. Master's thesis, University of Karlsruhe, 1994.

2) [Le Cun *et al.*, 1989] Y. Le Cun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard,W. Hubbard, and L. D. Jackel. Backpropogation applied to handwritten zip code recognition. *Neural Computation*, 1:541–551, 1989.

3) [Rowley *et al.*, 1995] Henry A. Rowley, Shumeet Baluja, and Takeo Kanade. Human face detection in visual scenes. CMUCS95158R, Carnegie Mellon University, November 1995.

4) [Sung and Poggio, 1994] KahKay Sung and Tomaso Poggio. Examplebased learning for viewbased human face detection. A.I. Memo 1521, CBCL Paper 112, MIT, December 1994.

5) [Umezaki, 1995] Tazio Umezaki. Personal communication, 1995.

7) [Vaillant *et al.*, 1994] R. Vaillant, C. Monrocq, and Y. Le Cun. Original approach for the localisation of objects in images. *IEE Proceedings on Vision, Image, and Signal Processing*, 141(4), August 1994.

8) [Waibel *et al.*, 1989] Alex Waibel, Toshiyuki Hanazawa, Geoffrey Hinton, Kiyohiro Shikano, and Kevin J. Lang. Phoneme recognition using timedelay neural networks. *Readings in Speech Recognition*, pages 393–404, 1989.