Automated Sperm Assessment Framework and Neural Network Specialized for Sperm Video Recognition

Takuro Fujii[1] Hayato Nakagawa[1] Teppei Takeshima[2] Yasushi Yumura[2] Tomoki Hamagami[1]

[1]Yokohama National University [2]Yokohama City University Medical Center

{tkr.fujii.ynu, haya.nakagawa.ynu, hamagami.ynu}@gmail.com {yumura, teppei_t}@yokohama-cu.ac.jp

A. Appendix

A.1. Domain Knowledge

In this section, we briefly introduce domain knowledge regarding sperm assessment and assisted reproductive technologies, according to the WHO semen analysis manual [1].

A.1.1 Biological Structure of Sperm

Structurally, a normal human sperm has three main parts: the head, neck, and tail, as shown in Figure 1. The tip of the sperm head contains acrosome that secretes enzymes that are useful for penetration, making the penetration process easier. The neck contains the mitochondria that supply the tail with energy for movement. The tail provides the sperm with motility for movement to oocytes for fertilization.

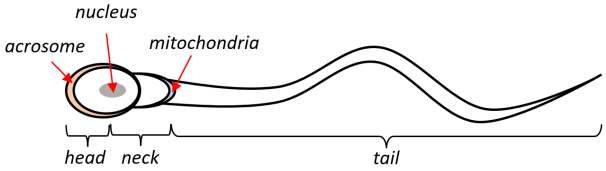

Figure 1. Diagram of human sperm. Normal human sperm has 3 main parts: head, neck and tail.

A.1.2 Assessment as Gold Standard

We present four key points that experts consider when assessing sperm.

1. **Head Shape:** The head should be smooth, regularly contoured, and generally oval shaped. The ratio of the head width to length should be approximately $3:5$.
2. **Acrosome Size:** The acrosome should comprise 40-70% of the head area.
3. **No Vacuoles:** The acrosome should contain no vacuoles and no more than two small vacuoles that should not occupy more than 20% of the sperm head.
4. **Neck:** The neck should be slender, regular, and approximately equal in length to the sperm head.

A.1.3 Assisted Reproductive Technologies (ARTs)

ARTs are primarily used to address infertility. ARTs includes in-vitro-fertilization (IVF), intracytoplasmic sperm injection (ICSI), cryopreservation of gametes or embryos, and the use of fertility drugs. See the WHO Semen Analysis Manual [1] for details on ARTs.

A.2. More Details on Our Paper

A.2.1 Samples of the Dataset

Some samples of the constructed data are shown in Figure 2. Original videos (left) are videos taken from a microscope and red regions surround target sperms.

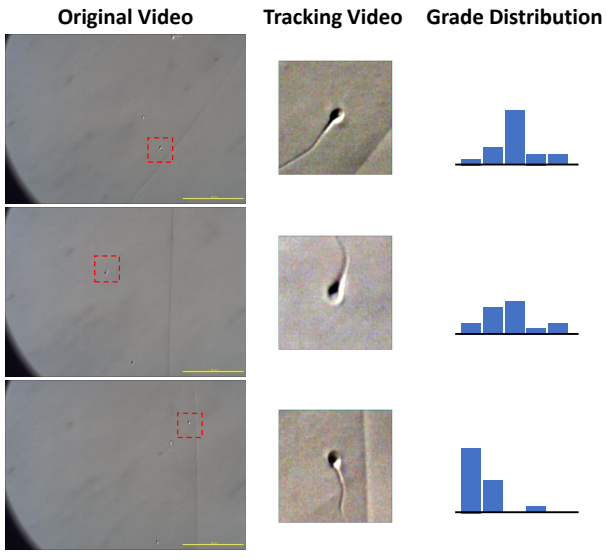

Figure 2. Some samples of the dataset: Original videos (left), Trancking Video (center) and Grade Distribution (right).

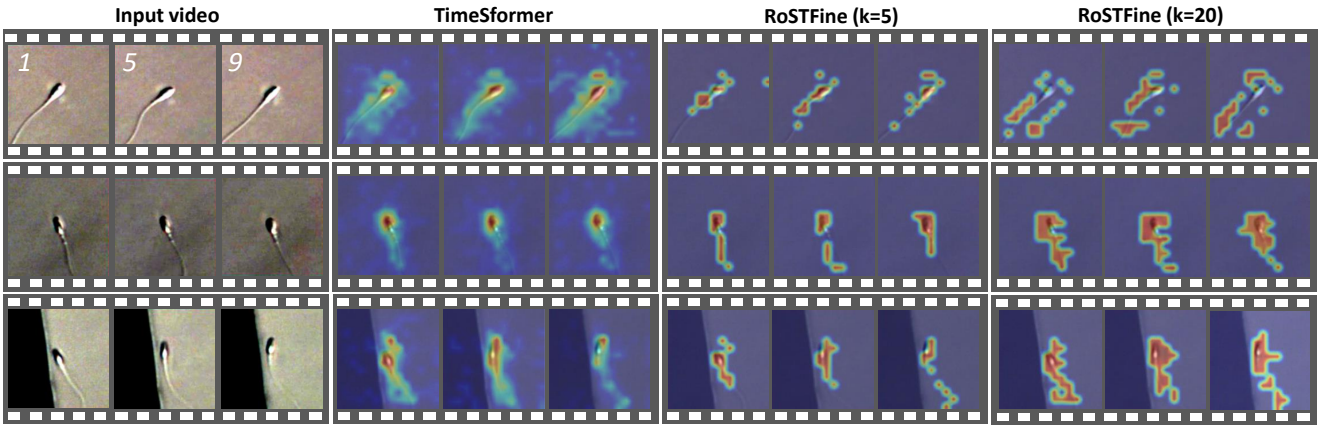

Figure 3. Visualization of attention maps of TimeSformer and RoSTFine ($K = 5, 20$).

A.3. TimeSformer v.s. RoSTFine

Table 1 shows performances on each fold. RoSTFine$_{\alpha=0}$ outperforms TimeSformer in 4 folds out of 5-fold cross validation in MSE and JS-divergence.

Fold	**TimeSformer**		RoSTFine$_{\alpha=0}$	
	MSE$^{(10^{-2})}$	JS$^{(10^{-2})}$	MSE$^{(10^{-2})}$	JS$^{(10^{-2})}$
1th	1.138	**4.097**	**1.090**	4.184
2th	1.112	4.293	**1.067**	**4.158**
3th	1.371	4.560	**1.340**	**4.558**
4th	1.190	4.356	**1.073**	**4.166**
5th	**1.032**	4.111	**1.032**	**3.907**

Table 1. Each fold performance in TimeSformer and RoSTFine. RoSTFine outperforms TimeSformer in 4 folds out of 5-fold cross validation in the MSE and JS-divergence.

A.4. Visualizing Attention Map

In this section, we present the attention maps of TimeSformer and RoSTFine in addition to those in §5.3. In Figure 3, we present space attention visualizations of 1st, 5th and 9th frames obtained using TimeSformer and RoSTFine($K = 5, 20$). The visualization method is the same as that described in §5.3.

In the case of $K = 5$, we observe that PSM enables the reduction of redundancy and attends more strongly to the sperm in Figure3. Although TimeSformer's attentions are vaguely oriented around the sperm, RoSTFine's attentions are strongly oriented only to the sperm head, neck and tail of the sperm. This results are the same as those of §5.3. However, in the case of $K = 20$, RoSTFine captures the sperm as well as a little background.

References

[1] World Health Organization. *WHO laboratory manual for the examination and processing of human semen*. World Health Organization, 2021.